STATISTICAL METHODS

Tom A. B. Snijders

http://www.stats.ox.ac.uk/~snijders/sm.htm

Lectures on Multilevel Analysis

Department of Statistics
University of Oxford

2012

This is a set of slides following Snijders & Bosker (2012).

The page headings give the chapter numbers and the page numbers in the book.

Literature:

Tom Snijders & Roel Bosker,
Multilevel Analysis: An Introduction to Basic and Applied Multilevel Analysis,
2**nd** edition. Sage, 2012.
Chapters 1-2, 4-6, 8, 10.

There is an associated website
http://www.stats.ox.ac.uk/~snijders/mlbook.htm
containing data sets and scripts for R and other software.

These slides are *not* self-contained, for understanding them it is necessary
also to study the corresponding parts of the book, and the R scripts at the website!

If you wish to see further literature, look at:

Andrew Gelman & Jennifer Hill,
Data Analysis Using Regression and Multilevel/Hierarchical Models. CUP, 2007.

For doing multilevel analysis using R, here are some R materials:

José Pinheiro & Douglas Bates,
Mixed-effects models in S and S-PLUS. Springer, 2000.

John Fox, *Linear Mixed Models. Appendix to 'An R and S-PLUS Companion to Applied Regression'.*

http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-mixed-models.pdf

Douglas Bates, *Examples from Multilevel Software Comparative Reviews*.
http://finzi.psych.upenn.edu/R/library/mlmRev/doc/MlmSoftRev.pdf

For further R literature see Section 18.2.2 of Snijders & Bosker.

2. Multilevel data and multilevel analysis

Multilevel Analysis using the hierarchical linear model :
random coefficient regression analysis for data with several nested levels.

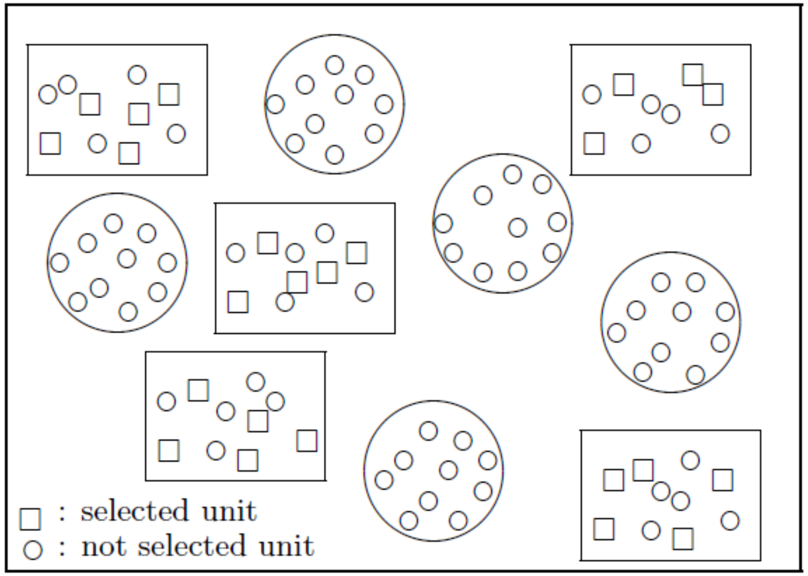


Figure 2.1: Multi-stage sample.

Each level is (potentially) a *source of unexplained variability*.

Some examples of units
at the macro and micro level:

macro-level	micro-level
schools	teachers
classes	pupils
neighborhoods	families
districts	voters
firms	departments
departments	employees
families	children
litters	animals
doctors	patients
interviewers	respondents
judges	suspects
subjects	measurements
respondents = egos	alters

Multilevel analysis is a suitable approach to take into account the *social contexts* as well as the *individual respondents* or *subjects*.

The hierarchical linear model is a type of regression analysis for multilevel data where the dependent variable is at the lowest level.

Explanatory variables can be defined at any level
(including aggregates of level-one variables).

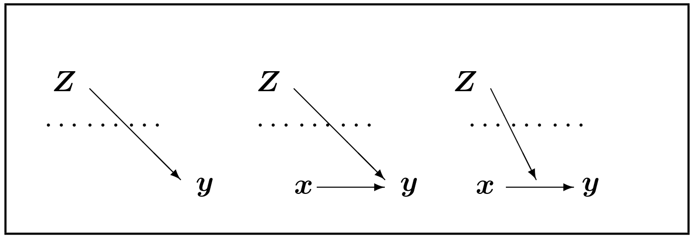


Figure 2.5 The structure of macro–micro propositions.

Also longitudinal data can be regarded as a nested structure;
for such data the hierarchical linear model is likewise convenient.

Two kinds of argument to choose for a multilevel analysis instead of an OLS regression of disaggregated data:

1. *Dependence as a nuisance*
   Standard errors and tests base on OLS regression are suspect
   because the assumption of independent residuals is invalid.

2. *Dependence as an interesting phenomenon*
   It is interesting in itself to disentangle variability at the various levels;
   moreover, this can give insight in where further explanation may fruitfully be sought.

4. The random intercept model

Hierarchical Linear Model:

$\boldsymbol{i}$ indicates level-one unit (e.g., individual);
$\boldsymbol{j}$ indicates level-two unit (e.g., group).

Variables for individual $\boldsymbol{i}$ in group $\boldsymbol{j}$:

$\boldsymbol{Y_{ij}}$ dependent variable;
$\boldsymbol{x_{ij}}$ explanatory variable at level one;

for group $\boldsymbol{j}$:

$\boldsymbol{z_j}$ explanatory variable at level two; $\boldsymbol{n_j}$ group size.

OLS regression model of $\boldsymbol{Y}$ on $\boldsymbol{X}$ ignoring groups :

$$\boldsymbol{Y_{ij} = \beta_0 + \beta_1 \, x_{ij} + R_{ij}} \; .$$

Group-dependent regressions:

$$\boldsymbol{Y_{ij} = \beta_{0j} + \beta_{1j} \, x_{ij} + R_{ij}} \; .$$

Distinguish two kinds of *fixed effects* models:

1. models where group structure is ignored;
2. models with fixed effects for groups: β_{0j} are fixed parameters.

In the *random intercept* model, the intercepts β_{0j} are random variables representing random differences between groups:

$$Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + R_{ij} \ .$$

where β_{0j} = average intercept γ_{00} plus group-dependent deviation U_{0j} :

$$\beta_{0j} = \gamma_{00} + U_{0j} \ .$$

In this model, the regression coefficient β_1 is common to all the groups.

In the random intercept model, the constant regression coefficient β_1 is sometimes denoted γ_{10}:

Substitution yields

$$Y_{ij} = \gamma_{00} + \gamma_{10}\, x_{ij} + U_{0j} + R_{ij} \; .$$

In the hierarchical linear model, the U_{0j} are *random* variables
and the statistical parameter in the model is not their individual values, but their variance

$$\tau^2 = \text{var}(U_{0j}).$$

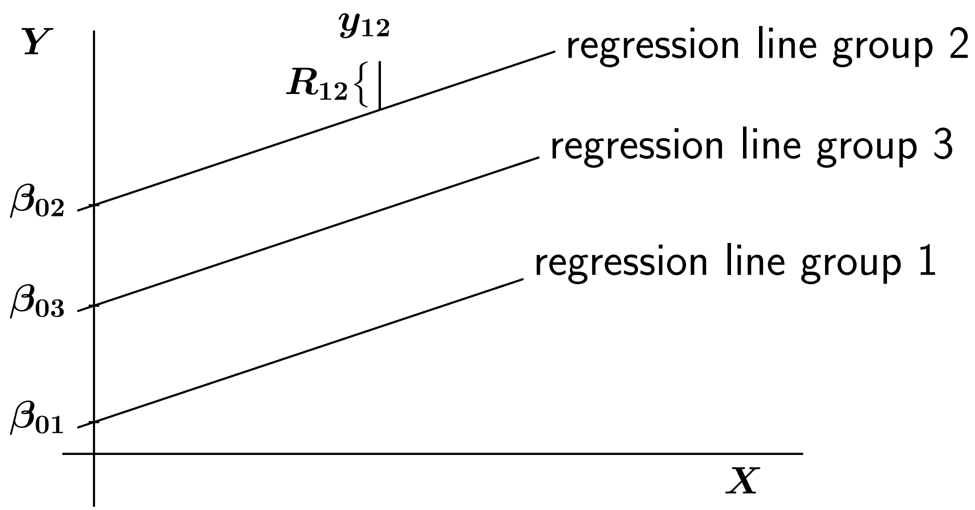


Figure 4.1 Different parallel regression lines.

The point y_{12} is indicated with its residual R_{12} .

Arguments for choosing between fixed (F) and random (R) coefficient models for the group dummies:

1. If groups are unique entities and inference should focus on *these* groups: F .
   This often is the case with a small number of groups.
2. If groups are regarded as sample from a (perhaps hypothetical) population and inference should focus on this population, then R .
   This often is the case with a large number of groups.
3. If level-two effects are to be tested, then R .
4. If group sizes are small and there are many groups, and it is reasonable to assume exchangeability of group-level residuals, then R makes better use of the data.
5. If the researcher is interested only in *within-group* effects, and is suspicious about the model for *between-group* differences, then F is more robust.
6. If group effects $\boldsymbol{U_{0j}}$ (etc.) are not nearly normally distributed, R is risky
   (or use more complicated multilevel models).

The empty model (*random effects ANOVA*) is a model
without explanatory variables:

$$\boldsymbol{Y_{ij}} = \gamma_{00} + \boldsymbol{U_{0j}} + \boldsymbol{R_{ij}} \; .$$

Variance decomposition:

$$\text{var}(\boldsymbol{Y_{ij}}) = \text{var}(\boldsymbol{U_{0j}}) + \text{var}(\boldsymbol{R_{ij}}) = \boldsymbol{\tau_0^2} + \boldsymbol{\sigma^2} \; .$$

Covariance between two individuals ($\boldsymbol{i \neq i'}$) in the same group $\boldsymbol{j}$:

$$\text{cov}(\boldsymbol{Y_{ij}}, \boldsymbol{Y_{i'j}}) = \text{var}(\boldsymbol{U_{0j}}) = \boldsymbol{\tau_0^2} \; ,$$

and their correlation:

$$\boldsymbol{\rho(Y_{ij}, Y_{i'j}) = \rho_I(Y) = \frac{\tau_0^2}{(\tau_0^2 + \sigma^2)}} \; .$$

This is the *intraclass correlation coefficient*.

Often between .05 and .25 in social science research,
where the groups represent some kind of social grouping.

Example: 3758 pupils in 211 schools , Y = language test.

Classrooms / schools are level-2 units.

Table 4.1 Estimates for empty model

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.00	0.32
Random Part	**Variance Component**	**S.E.**
Level-two variance:		
$\tau_0^2 = \text{var}(U_{0j})$	18.12	2.16
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	62.85	1.49
Deviance	26595.3	

Intraclass correlation

$$\rho_{\mathrm{I}} = \frac{18.12}{18.12 + 62.85} = 0.22$$

Total population of individual values Y_{ij} has estimated mean 41.00 and standard deviation $\sqrt{18.12 + 62.85} = 9.00$.

Population of class means β_{0j} has estimated mean 41.00 and standard deviation $\sqrt{18.12} = 4.3$.

The model becomes more interesting,
when also *fixed effects* of explanatory variables are included:

$$Y_{ij} = \gamma_{00} + \gamma_{10}\, x_{ij} + U_{0j} + R_{ij} \; .$$

(Note the difference between fixed effects of explanatory variables and fixed effects of group dummies!)

Table 4.2 Estimates for random intercept model with effect for IQ

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.06	0.24
γ_{10} = Coefficient of IQ	2.507	0.054

Random Part	Variance Component	S.E.
Level-two variance:		
$\boldsymbol{\tau_0^2}$ = var($\boldsymbol{U_{0j}}$)	9.85	1.21
Level-one variance:		
$\boldsymbol{\sigma^2}$ = var($\boldsymbol{R_{ij}}$)	40.47	0.96
Deviance	24912.2	

There are two kinds of parameters:

1. fixed effects: regression coefficients γ (just like in OLS regression);
2. random effects: variance components $\boldsymbol{\sigma^2}$ and $\boldsymbol{\tau_0^2}$.

Table 4.3 Estimates for ordinary least squares regression

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.30	0.12
γ_{10} = Coefficient of IQ	2.651	0.056

Random Part	Variance Component	S.E.
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	49.80	1.15
Deviance	25351.0	

Multilevel model has more structure ("dependence interesting");

OLS has misleading standard error for intercept ("dependence nuisance").

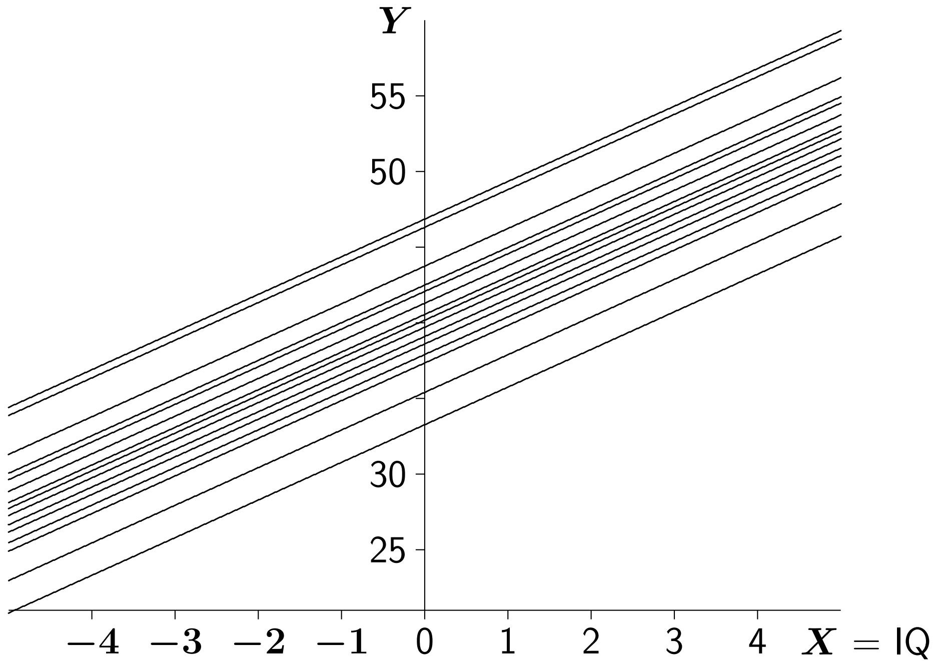


Figure 4.2 Fifteen randomly chosen regression lines according to the random intercept model of Table 4.2.

More explanatory variables:

$$\begin{aligned} \boldsymbol{Y}_{ij} = \gamma_{00} &+ \gamma_{10}\,\boldsymbol{x}_{1ij} + \ldots + \gamma_{p0}\,\boldsymbol{x}_{pij} + \gamma_{01}\,\boldsymbol{z}_{1j} + \ldots + \gamma_{0q}\,\boldsymbol{z}_{qj} \\ &+ \boldsymbol{U}_{0j} + \boldsymbol{R}_{ij}\,. \end{aligned}$$

Especially important:
difference between within-group and between-group regressions.

The within-group regression coefficient is the regression coefficient within each group, assumed to be the same across the groups.

The between-group regression coefficient is defined as the regression coefficient for the regression of the group means of $\boldsymbol{Y}$ on the group means of $\boldsymbol{X}$.

This distinction is essential to avoid *ecological fallacies* (p. 15–17 in the book).

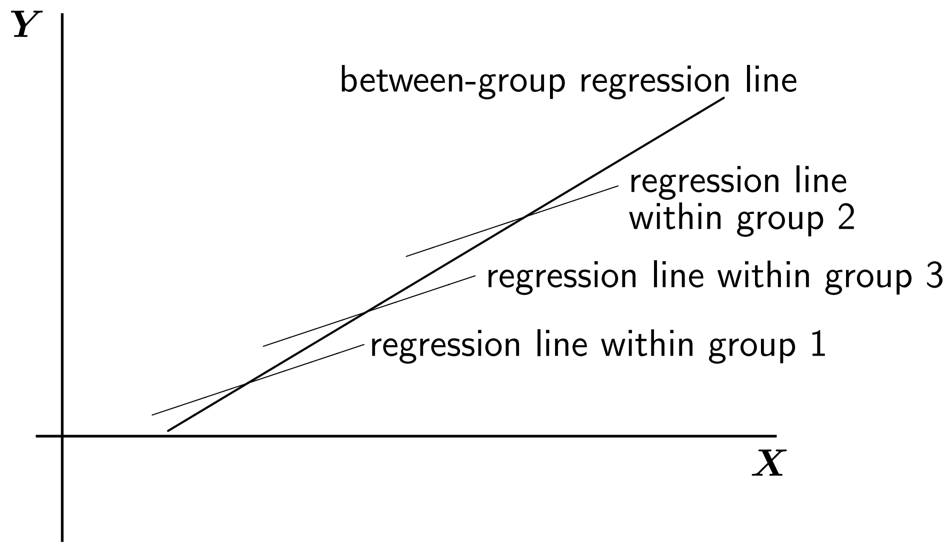


Figure 4.3 Different between-group and within-group regression lines.

This is obtained by having *separate fixed effects* for the level-1 variable $\boldsymbol{X}$ and its group mean $\bar{\boldsymbol{X}}$.
(Alternative:
use the within-group deviation variable $\tilde{\boldsymbol{X}}_{ij} = (\boldsymbol{X} - \bar{\boldsymbol{X}})$ instead of $\boldsymbol{X}$.)

Table 4.4 Estimates for random intercept model
with different within- and between-group regressions

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.11	0.23
γ_{10} = Coefficient of IQ	2.454	0.055
γ_{01} = Coefficient of $\overline{\text{IQ}}$ (group mean)	1.312	0.262

Random Part	Variance Component	S.E.
Level-two variance:		
$\tau_0^2 = \text{var}(U_{0j})$	8.68	1.10
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	40.43	0.96
Deviance	24888.0	

In the model with separate effects for the original variable x_{ij} and the group mean

$$Y_{ij} = \gamma_{00} + \gamma_{10}\, x_{ij} + \gamma_{01} \overline{x}_{.j} + U_{0j} + R_{ij} \ ,$$

the within-group regression coefficient is γ_{10} ,
between-group regression coefficient is $\gamma_{10} + \gamma_{01}$.
This is convenient because the difference between within-group and between-group coefficients can be tested by considering γ_{01}.

In the model with separate effects for group-centered variable $\tilde{x}_{ij}$
and the group mean

$$Y_{ij} = \tilde{\gamma}_{00} + \tilde{\gamma}_{10}\, \tilde{x}_{ij} + \tilde{\gamma}_{01} \overline{x}_{.j} + U_{0j} + R_{ij} \ ,$$

the within-group regression coefficient is $\tilde{\gamma}_{10}$,
the between-group regression coefficient is $\tilde{\gamma}_{01}$.
This is convenient because these coefficients are given immediately in the results, with their standard errors.

Both models are equivalent, and have the same fit: $\tilde{\gamma}_{10} = \gamma_{10}$, $\tilde{\gamma}_{01} = \gamma_{10} + \gamma_{01}$.

Estimation/prediction of random effects

The random effects U_{0j} are *not* statistical parameters and therefore they are not estimated as part of the estimation routine.

However, it sometimes is desirable to 'estimate' them. This can be done by the *empirical Bayes* method; these 'estimates' are also called the *posterior means*. In statistical terminology, this is not called 'estimation' but 'prediction', the name for the construction of likely values for unobserved random variables.

The posterior mean for group j is based on two kinds of information:

⇒ *sample information* : the data in group j;
⇒ *population information* : the value U_{0j} was drawn from a normal distribution with mean 0 and variance τ_0^2 .

If the population information is reasonable, this gives on average an improved prediction.

The empirical Bayes estimate in the case of the empty model is a weighted average of the group mean and the overall mean:

$$\hat{\beta}_{0j}^{\mathsf{EB}} = \lambda_j \, \hat{\beta}_{0j} + (1 - \lambda_j) \, \hat{\gamma}_{00} \ ,$$

where the weight λ_j is the 'reliability' of the mean of group j

$$\lambda_j = \frac{\tau_0^2}{\tau_0^2 + \sigma^2 / n_j} \ .$$

These 'estimates' are not unbiased for each specific group, but they are more precise when the mean squared errors are averaged over all groups.

For models with explanatory variables, the same principle can be applied: the values that would be obtained as OLS estimates per group are "shrunk towards the mean".

There are two kinds of standard errors for empirical Bayes estimates:

comparative standard errors

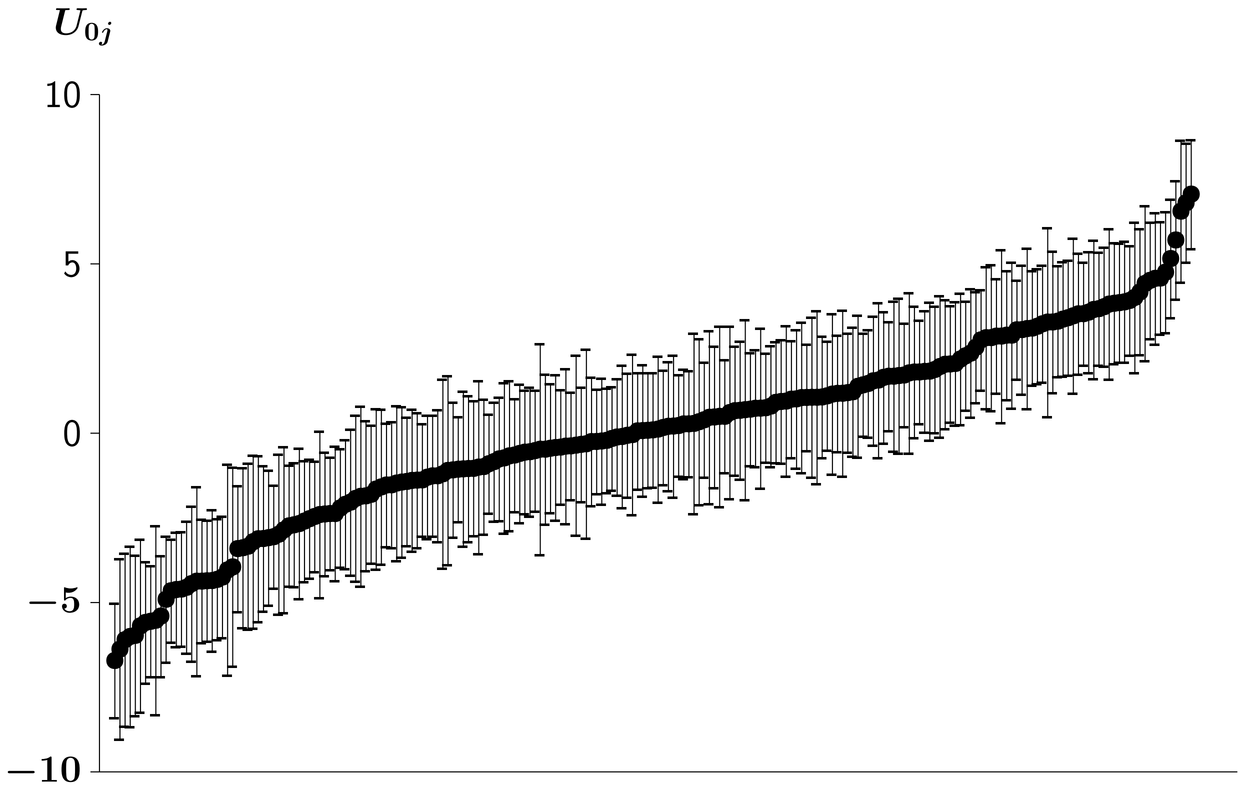
$$\text{S.E.}_{\text{comp}}\left(\hat{\boldsymbol{U}}^{\text{EB}}_{\boldsymbol{hj}}\right) = \text{S.E.}\left(\hat{\boldsymbol{U}}^{\text{EB}}_{\boldsymbol{hj}} - \boldsymbol{U}_{\boldsymbol{hj}}\right)$$

for comparing the random effects of different level-2 units
(use with caution – E.B. estimates are not unbiased!);

and *diagnostic standard errors*

$$\text{S.E.}_{\text{diag}}\left(\hat{\boldsymbol{U}}^{\text{EB}}_{\boldsymbol{hj}}\right) = \text{S.E.}\left(\hat{\boldsymbol{U}}^{\text{EB}}_{\boldsymbol{hj}}\right)$$

used for model checking (e.g., checking normality of the level-two residuals).



The ordered added value scores for 211 schools with comparative posterior confidence intervals.

In this figure, the error bars extend 1.39 times the comparative standard errors to either side, so that schools may be deemed to be significantly different if the intervals do not overlap (no correction for multiple testing!).

5. The hierarchical linear model

It is possible that not only the group average of $\boldsymbol{Y}$,
but also the effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is *randomly* dependent on the group.

In other words, in the equation

$$\boldsymbol{Y}_{ij} = \boldsymbol{\beta}_{0j} + \boldsymbol{\beta}_{1j}\, \boldsymbol{x}_{ij} + \boldsymbol{R}_{ij} ,$$

also the regression coefficient $\boldsymbol{\beta}_{1j}$ has a random part:

$$\begin{aligned} \boldsymbol{\beta}_{0j} &= \gamma_{00} + \boldsymbol{U}_{0j} \\ \boldsymbol{\beta}_{1j} &= \gamma_{10} + \boldsymbol{U}_{1j} . \end{aligned}$$

Substitution leads to

$$\boldsymbol{Y}_{ij} = \gamma_{00} + \gamma_{10}\, \boldsymbol{x}_{ij} + \boldsymbol{U}_{0j} + \boldsymbol{U}_{1j}\, \boldsymbol{x}_{ij} + \boldsymbol{R}_{ij} .$$

Variable $\boldsymbol{X}$ now has a *random slope*.

Again the group-dependent coefficients U_{0j}, U_{1j} are not individual parameters in the statistical sense, but only their variances, and covariance, are:

$$\begin{aligned} \text{var}(U_{0j}) &= \tau_{00} = \tau_0^2 \ ; \\ \text{var}(U_{1j}) &= \tau_{11} = \tau_1^2 \ ; \\ \text{cov}(U_{0j}, U_{1j}) &= \tau_{01} \ . \end{aligned}$$

Thus we have a linear model for the mean structure, and a parametrized covariance matrix within groups with independence between groups.

5.1 Estimates for random slope model

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.127	0.234
γ_{10} = Coeff.	2.480	0.064
γ_{01} = Coeff. of $\overline{\text{IQ}}$ (group mean)	1.029	0.262

Random Part	Parameters	S.E.
Level-two random part:		
$\tau_0^2 = \text{var}(U_{0j})$	8.877	1.117
$\tau_1^2 = \text{var}(U_{1j})$	0.195	0.076
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	-0.835	0.217
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	39.685	0.964
Deviance	24864.9	

$\overline{\text{IQ}}$ is defined as the group mean.

The equation for this table is

$$\begin{aligned} Y_{ij} \;=\; & 41.13 \;+\; 2.480\,\text{IQ}_{ij} \\ & + 1.029\,\overline{\text{IQ}}_{.j} \\ & + U_{0j} \;+\; U_{1j}\,\text{IQ}_{ij} \;+\; R_{ij}\ . \end{aligned}$$

The slope β_{1j} has average 2.480 and s.d. $\sqrt{0.195} = 0.44$.

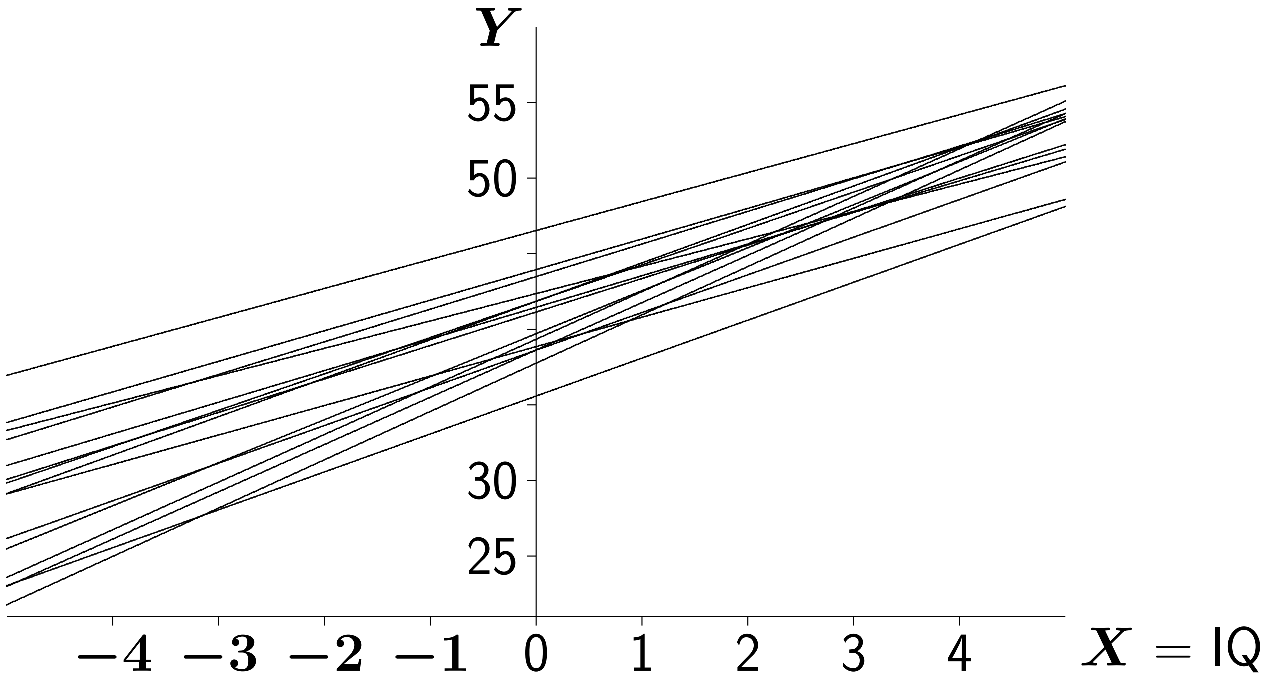


Figure 5.2 Fifteen random regression lines according to the model of Table 5.1.

Note the heteroscedasticity: variance is larger for low X than for high X.
The lines fan in towards the right.

Intercept variance and intercept-slope covariance depend on the position of the $X = 0$ value, because the intercept is defined by the $X = 0$ axis.

The next step is to *explain* the random slopes:

$$\boldsymbol{\beta_{0j}} = \gamma_{00} + \gamma_{01}\, \boldsymbol{z_j} + \boldsymbol{U_{0j}}$$

$$\boldsymbol{\beta_{1j}} = \gamma_{10} + \gamma_{11}\, \boldsymbol{z_j} + \boldsymbol{U_{1j}} \ .$$

Substitution then yields

$$\begin{aligned} \boldsymbol{Y_{ij}} &= (\gamma_{00} + \gamma_{01}\, \boldsymbol{z_j} + \boldsymbol{U_{0j}}) \\ &\quad + (\gamma_{10} + \gamma_{11}\, \boldsymbol{z_j} + \boldsymbol{U_{1j}})\, \boldsymbol{x_{ij}} + \boldsymbol{R_{ij}} \\ &= \gamma_{00} + \gamma_{01}\, \boldsymbol{z_j} + \gamma_{10}\, \boldsymbol{x_{ij}} + \gamma_{11}\, \boldsymbol{z_j}\, \boldsymbol{x_{ij}} \\ &\qquad + \boldsymbol{U_{0j}} + \boldsymbol{U_{1j}}\, \boldsymbol{x_{ij}} + \boldsymbol{R_{ij}} \ . \end{aligned}$$

The term $\gamma_{11}\, \boldsymbol{z_j}\, \boldsymbol{x_{ij}}$ is called the *cross-level interaction effect*.

Table 5.2 Estimates for model with random slope and cross-level interaction

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.254	0.235
γ_{10} = Coefficient of IQ	2.463	0.063
γ_{01} = Coefficient of $\overline{\text{IQ}}$	1.131	0.262
γ_{11} = Coefficient of $\overline{\text{IQ}} \times$ IQ	-0.187	0.064

Random Part	Parameters	S.E.
Level-two random part:		
$\tau_0^2 = \text{var}(U_{0j})$	8.601	1.088
$\tau_1^2 = \text{var}(U_{1j})$	0.163	0.072
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	-0.833	0.210
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	39.758	0.965
Deviance	24856.8	

For two variables (IQ and SES) and two levels (student and school), the main effects and interactions give rise to a lot of possible combinations:

Table 5.3 Estimates for model with random slopes and many effects

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.632	0.255
γ_{10} = Coefficient of IQ	2.230	0.063
γ_{20} = Coefficient of SES	0.172	0.012
γ_{30} = Interaction of IQ and SES	–0.019	0.006
γ_{01} = Coefficient of $\overline{\text{IQ}}$	0.816	0.308
γ_{02} = Coefficient of $\overline{\text{SES}}$	–0.090	0.044
γ_{03} = Interaction of $\overline{\text{IQ}}$ and $\overline{\text{SES}}$	–0.134	0.037
γ_{11} = Interaction of IQ and $\overline{\text{IQ}}$	–0.081	0.081
γ_{12} = Interaction of IQ and $\overline{\text{SES}}$	0.004	0.013
γ_{21} = Interaction of SES and $\overline{\text{IQ}}$	0.023	0.018
γ_{22} = Interaction of SES and $\overline{\text{SES}}$	0.000	0.002

(continued next page....)

Random Part	Parameters	S.E.
Level-two random part:		
$\tau_0^2 = \text{var}(U_{0j})$	8.344	1.407
$\tau_1^2 = \text{var}(U_{1j})$	0.165	0.069
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	–0.942	0.204
$\tau_2^2 = \text{var}(U_{2j})$	0.0	0.0
$\tau_{02} = \text{cov}(U_{0j}, U_{2j})$	0.0	0.0
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	37.358	0.907
Deviance	24624.0	

The non-significant parts of the model may be dropped:

Table 5.4 Estimates for a more parsimonious model with a random slope and many effects

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.612	0.247
γ_{10} = Coefficient of IQ	2.231	0.063
γ_{20} = Coefficient of SES	0.174	0.012
γ_{30} = Interaction of IQ and SES	–0.017	0.005
γ_{01} = Coefficient of $\overline{\text{IQ}}$	0.760	0.296
γ_{02} = Coefficient of $\overline{\text{SES}}$	–0.089	0.042
γ_{03} = Interaction of $\overline{\text{IQ}}$ and $\overline{\text{SES}}$	–0.120	0.033

Random Part	Parameters	S.E.
Level-two random part:		
$\tau_0^2 = \text{var}(U_{0j})$	8.369	1.050
$\tau_1^2 = \text{var}(U_{1j})$	0.164	0.069
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	–0.929	0.204
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	37.378	0.907
Deviance	24626.8	

General formulation of the two-level model

As a link to the general statistical literature,
it may be noted that the two-level model can be expressed as follows:

$$\boldsymbol{Y_j} = \boldsymbol{X_j}\,\gamma \;+\; \boldsymbol{Z_j U_j} \;+\; \boldsymbol{R_j}$$

with $\begin{bmatrix} \boldsymbol{R_j} \\ \boldsymbol{U_j} \end{bmatrix} \sim \boldsymbol{\mathcal{N}} \left(\begin{bmatrix} \emptyset \\ \emptyset \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma_j(\theta)} & \emptyset \\ \emptyset & \boldsymbol{\Omega(\xi)} \end{bmatrix} \right)$

and $(\boldsymbol{R_j}, \boldsymbol{U_j}) \perp (\boldsymbol{R_\ell}, \boldsymbol{U_\ell})$ for all $\boldsymbol{j \neq \ell}$.

Standard specification $\boldsymbol{\Sigma_j(\theta) = \sigma^2 I_{n_j}}$,
but other specifications are possible.
Mostly, $\boldsymbol{\Sigma_j(\theta)}$ is diagonal, but even this is not necessary (e.g. time series).

The model formulation yields

$$\boldsymbol{Y_j} \sim \mathcal{N}\left(\boldsymbol{X_j\gamma},\ \boldsymbol{Z_j\Omega(\xi)Z_j'} + \boldsymbol{\Sigma_j(\theta)}\right) .$$

This is a special case of the *mixed linear model*

$$\boldsymbol{Y} = \boldsymbol{X\gamma} + \boldsymbol{ZU} + \boldsymbol{R},$$

with $\boldsymbol{X}[\boldsymbol{n},\boldsymbol{r}]$, $\boldsymbol{Z}[\boldsymbol{n},\boldsymbol{p}]$, and

$$\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{U}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\emptyset\\ \emptyset\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma} & \emptyset\\ \emptyset & \boldsymbol{\Omega}\end{pmatrix}\right).$$

For estimation, the ML and REML methods are mostly used.

These can be implemented by various algorithms: Fisher scoring,
EM = Expectation–Maximization, IGLS = Iterative Generalized Least Squares.
See Section 4.7 and 5.4.
This is not examinable material.

Level-1 heteroscedasticity (see Chapter 8)

The following formulation allows for heteroscedasticity depending linearly/quadratically on level-1 variables $\boldsymbol{V}$:

$$\boldsymbol{R}_j = \begin{bmatrix} \boldsymbol{R}_{1j} \\ \dots \\ \boldsymbol{R}_{n_j j} \end{bmatrix} \text{ with } \boldsymbol{R}_{ij} = \boldsymbol{v}_{ij}\, \boldsymbol{R}^0_{ij}$$

where

$\boldsymbol{v}_{ij}$ is a $\mathbf{1} \times \boldsymbol{t}$ variable ,

$\boldsymbol{R}^0_{ij}$ is a $\boldsymbol{t} \times \mathbf{1}$ random vector ,

$$\boldsymbol{R}^0_{ij} \sim \mathcal{N}(0, \Sigma^0(\boldsymbol{\theta})) \ .$$

This implies

$$\textit{Var}\ \boldsymbol{R}_{ij} = \boldsymbol{v}_{ij} \boldsymbol{\Sigma}^0(\boldsymbol{\theta}) \boldsymbol{v}'_{ij} \ .$$

It does not matter if $\boldsymbol{\Sigma}^0(\boldsymbol{\theta})$ is not positive semi-definite, as long as the resulting *Var* $\boldsymbol{R}_{ij}$ is p.s.d.

E.g., linear variance function for

$$\boldsymbol{\Sigma}^0(\boldsymbol{\theta}) = (\boldsymbol{\sigma}_{hk}(\boldsymbol{\theta}))_{1 \le h,k \le t}$$

is obtained with with

$$\begin{aligned} \boldsymbol{\sigma}_{h1}(\boldsymbol{\theta}) &= \boldsymbol{\sigma}_{1h}(\boldsymbol{\theta}) = \boldsymbol{\theta}_h \quad & h = 1, \ldots, t \\ \boldsymbol{\sigma}_{hk}(\boldsymbol{\theta}) &= 0 & \min\{h, k\} \ge 2 \;. \end{aligned}$$

More generally, any quadratic variance function can be obtained.

6. Testing

To test fixed effects, use the t-test with test statistic

$$T(\gamma_h) = \frac{\hat{\gamma}_h}{\text{S.E.}(\hat{\gamma}_h)} \ .$$

(Or the Wald test for testing several parameters simultaneously.)

For parameters in the random part, do not use t-tests.

Simplest test for any parameters (fixed and random parts) is the *deviance* (likelihood ratio) test, which can be used when comparing two model fits that have used the same set of cases: subtract deviances, use chi-squared test ($d.f.$ = number of parameters tested).

Other tests for parameters in the random part have been developed which are similar to F-tests in ANOVA.

6.1 Two models with different between- and within-group regressions

	Model 1		Model 2	
Fixed Effects	Coefficient	S.E.	Coefficient	S.E.
γ_{00} = Intercept	41.15	0.23	41.15	0.23
γ_{10} = Coeff. of IQ	2.265	0.065		
γ_{20} = Coeff. of $\widetilde{\text{IQ}}$			2.265	0.065
γ_{30} = Coeff. of SES	0.161	0.011	0.161	0.011
γ_{01} = Coeff. of $\overline{\text{IQ}}$	0.647	0.264	2.912	0.262
Random Part	Parameter	S.E.	Parameter	S.E.
Level-two parameters:				
$\tau_0^2 = \text{var}(U_{0j})$	9.08	1.12	9.08	1.12
$\tau_1^2 = \text{var}(U_{1j})$	0.197	0.074	0.197	0.074
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	-0.815	0.214	-0.815	0.214
Level-one variance:				
$\sigma^2 = \text{var}(R_{ij})$	37.42	0.91	37.42	0.91
Deviance	24661.3		24661.3	

Test for equality of within- and between-group regressions is t-test for $\overline{\text{IQ}}$ in Model 1:
$t = 0.647/0.264 = 2.45$,
$p < 0.02$.

Model 2 gives
within-group coefficient 2.265
and between-group coefficient
2.912 = 2.265 + 0.647.

However, one special circumstance: variance parameters are necessarily positive.
Therefore, they may be tested one-sided.

E.g., in the random intercept model
under the null hypothesis that $\tau_0^2 = 0$,
the asymptotic distribution of –2 times the log-likelihood ratio (deviance difference)
is a mixture of a point mass at 0 (with probability $\frac{1}{2}$)
and a χ^2 distribution (also with probability $\frac{1}{2}$.)

The interpretation is that if the observed between-group variance
is less than expected under the null hypothesis
– which happens with probability $\frac{1}{2}$ –
the estimate is $\hat{\tau}_0^2 = 0$ and the log-likelihood ratio is 0.

The test works as follows:
if deviance difference $= 0$, then no significance;
if deviance difference > 0, calculate p-value from χ_1^2 and divide by 2.

For testing random slope variances,
if the number of tested parameters (variances & covariances) is $p + 1$,
the p-values can be obtained as
the average of the p-values for the χ^2_p and χ^2_{p+1} distributions.

(Apologies for the use of the letter p in two different meanings...)

See p. 99.

Sections 6.3 and 6.4 are not treated in these slides.
You are requested to study them so that you understand the reasoning.
Details will not be examined,
but it is expected that you can apply this type of arguments.

8. Heteroscedasticity

The multilevel model allows to formulate heteroscedastic models where residual variance depends on observed variables.

E.g., random part at level one = $\boldsymbol{R_{0ij} + R_{1ij}\, x_{1ij}}$.

Then the level-1 variance is a quadratic function of $\boldsymbol{X}$:

var$\boldsymbol{(R_{0ij} + R_{1ij}\, x_{ij}) = \sigma_0^2 + 2\,\sigma_{01}\, x_{1ij} + \sigma_1^2\, x_{1ij}^2}$.

For $\boldsymbol{\sigma_1^2 = 0}$, this is a linear function:

var$\boldsymbol{(R_{0ij} + R_{1ij}\, x_{ij}) = \sigma_0^2 + 2\,\sigma_{01}\, x_{1ij}}$.

Possible as a variance function, without random effects interpretation.

8.1 Homoscedastic and heteroscedastic models.

	Model 1		Model 2	
Fixed Effect	Coefficient	S.E.	Coefficient	S.E.
Intercept	40.426	0.265	40.435	0.266
IQ	2.249	0.062	2.245	0.062
SES	0.171	0.011	0.171	0.011
IQ × SES	–0.020	0.005	–0.019	0.005
Gender	2.407	0.201	2.404	0.201
$\overline{\text{IQ}}$	0.769	0.293	0.749	0.292
$\overline{\text{SES}}$	–0.093	0.042	–0.091	0.042
$\overline{\text{IQ}} \times \overline{\text{SES}}$	–0.105	0.033	–0.107	0.033
Random Part	Parameters	S.E.	Parameters	S.E.
Level-two random part:				
Intercept variance	8.321	1.036	8.264	1.030
IQ slope variance	0.146	0.065	0.146	0.065
Intercept - IQ slope covariance	−0.898	0.197	−0.906	0.197
Level-one variance:				
σ_0^2 constant term	35.995	0.874	37.851	1.280
σ_{01} gender effect			–1.887	0.871
Deviance	24486.8		24482.2	

This shows that there is significant evidence for heteroscedasticity:
$\boldsymbol{\chi^2_1 = 4.6,\ p < 0.05}$.

The estimated residual (level-1) variance is
37.85 for boys and 37.85 – 2×1.89 = 34.07 for girls.

The following models show, however, that the heteroscedasticity as a function of IQ is more important.
First look only at Model 3.

8.2 Heteroscedastic models depending on IQ.

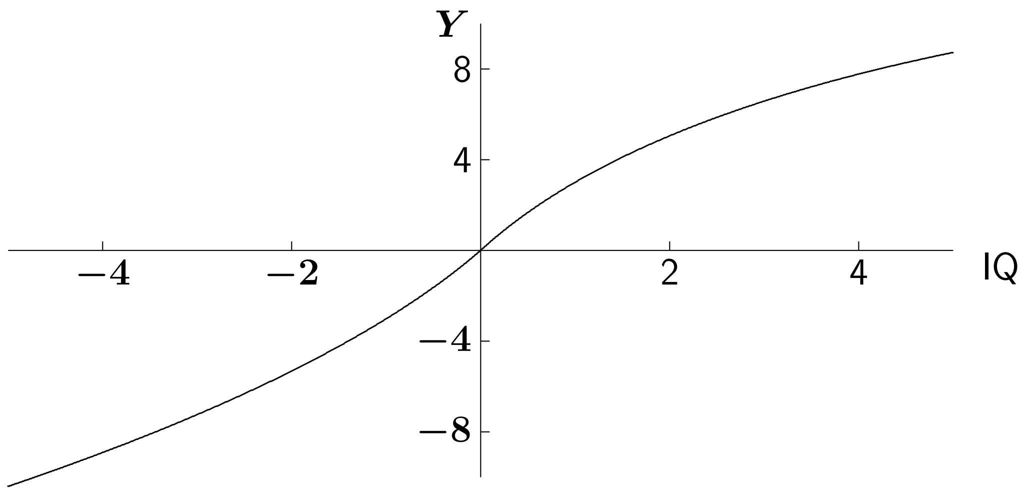
	Model 3		Model 4	
Fixed Effect	Coefficient	S.E.	Coefficient	S.E.
Intercept	40.51	0.26	40.51	0.27
IQ	2.200	0.058	3.046	0.125
SES	0.175	0.011	0.168	0.011
IQ × SES	–0.022	0.005	–0.016	0.005
Gender	2.311	0.198	2.252	0.196
$\overline{\text{IQ}}$	0.685	0.289	0.800	0.284
$\overline{\text{SES}}$	–0.087	0.041	–0.083	0.041
$\overline{\text{IQ}} \times \overline{\text{SES}}$	–0.107	0.033	–0.089	0.032
IQ^2_-			0.193	0.038
IQ^2_+			-0.260	0.033
Random Part	Parameter	S.E.	Parameter	S.E.
Level-two random effects:				
Intercept variance	8.208	1.029	7.989	1.002
IQ slope variance	0.108	0.057	0.044	0.048
Intercept - IQ slope covariance	–0.733	0.187	–0.678	0.171
Level-one variance parameters:				
σ_0^2 constant term	36.382	0.894	36.139	0.887
σ_{01} IQ effect	–1.689	0.200	–1.769	0.191
Deviance	24430.2		24369.0	

The level-1 variance function for Model 3 is 36.38 – 3.38 IQ .

Maybe further differentiation is possible between low-IQ pupils?
Model 4 uses

$$\text{IQ}^2_- = \begin{cases} \text{IQ}^2 & \text{if IQ } < 0 \\ 0 & \text{if IQ } \geq 0 , \end{cases}$$

$$\text{IQ}^2_+ = \begin{cases} 0 & \text{if IQ } < 0 \\ \text{IQ}^2 & \text{if IQ } \geq 0 . \end{cases}$$



Effect of IQ on language test as estimated by Model 4.

Heteroscedasticity can be very important for the researcher (although mostly she/he doesn't know it yet).

Bryk & Raudenbush: Correlates of diversity.
Explain not only means, but also variances!

Heteroscedasticity also possible for level-2 random effects:
give a random slope at level 2 to a level-2 variable.

10. Assumptions of the Hierarchical Linear Model

$$Y_{ij} = \gamma_0 + \sum_{h=1}^{r} \gamma_h x_{hij} + U_{0j} + \sum_{h=1}^{p} U_{hj} x_{hij} + R_{ij} \,.$$

Questions:

1. Does the fixed part contain the right variables (now X_1 to X_r)?
2. Does the random part contain the right variables (now X_1 to X_p)?
3. Are the level-one residuals normally distributed?
4. Do the level-one residuals have constant variance?
5. Are the level-two random coefficients normally distributed with mean 0?
6. Do the level-two random coefficients have a constant covariance matrix?

Follow the logic of the HLM

1. Include contextual effects

For every level-1 variable $\boldsymbol{X_h}$, check the fixed effect of the group mean $\bar{\boldsymbol{X}}_{\boldsymbol{h}}$.

Econometricians' wisdom: "the $\boldsymbol{U_{0j}}$ must not be correlated with the $\boldsymbol{X_{hij}}$.
Therefore test this correlation by testing the effect of $\bar{\boldsymbol{X}}_{\boldsymbol{h}}$ ('Hausman test')
Use a fixed effects model if this effect is significant".

Different approach to the same assumption:
Include the fixed effect of $\bar{\boldsymbol{X}}_{\boldsymbol{h}}$ if it is significant,
and continue to use a random effects model.
(Also check effects of variables $\bar{\boldsymbol{X}}_{\boldsymbol{h.j}}\, \boldsymbol{Z_j}$ for cross-level interactions involving $\boldsymbol{X_h}$!)

Also the random slopes $\boldsymbol{U_{hj}}$ must not be correlated with the $\boldsymbol{X_{kij}}$.
This can be checked by testing the fixed effect of $\bar{\boldsymbol{X}}_{\boldsymbol{k.j}}\, \boldsymbol{X_{hij}}$.

This procedure widens the scope of random coefficient models beyond what is allowed by the conventional rules of econometricians.

Assumption that level-2 random effects $\boldsymbol{U}_j$ have zero means.

What kind of bias can occur if this assumption is made but does not hold?

For a misspecified model,
suppose that we are considering a random intercept model:

$$\boldsymbol{Z}_j = \mathbf{1}_j$$

where the expected value of $\boldsymbol{U}_j$ is not 0 but

$$E\boldsymbol{U}_j = \boldsymbol{z}_{2j}\,\gamma_\star$$

for $\mathbf{1} \times \boldsymbol{r}$ vectors $\boldsymbol{z}_{2j}$ and an unknown regression coefficient $\boldsymbol{\gamma}_\star$. Then

$$\boldsymbol{U}_j = \boldsymbol{z}_{2j}\,\gamma_\star + \tilde{\boldsymbol{U}}_j$$

with

$$E\tilde{\boldsymbol{U}}_j = \mathbf{0}\ .$$

Write $\boldsymbol{X}_j = \bar{\boldsymbol{X}}_j + \tilde{\boldsymbol{X}}_j$, where $\bar{\boldsymbol{X}}_j = \mathbf{1}_j\,(\mathbf{1}_j'\mathbf{1}_j)^{-1}\mathbf{1}_j'\,\boldsymbol{X}_j$ are the group means.
Then the data generating mechanism is

$$\boldsymbol{Y}_j = \bar{\boldsymbol{X}}_j\,\gamma + \tilde{\boldsymbol{X}}_j\,\gamma + \mathbf{1}_j\,\boldsymbol{z}_{2j}\,\gamma_\star + \mathbf{1}_j\,\tilde{\boldsymbol{U}}_j + \boldsymbol{R}_j\ ,$$

where $E\tilde{\boldsymbol{U}}_j = \mathbf{0}$.
There will be a bias in the estimation of γ
if the matrices $\boldsymbol{X}_j = \bar{\boldsymbol{X}}_j + \tilde{\boldsymbol{X}}_j$ and $\mathbf{1}_j\,\tilde{\boldsymbol{U}}_j$ are not orthogonal.
By construction, $\tilde{\boldsymbol{X}}_j$ and $\mathbf{1}_j\,\tilde{\boldsymbol{U}}_j$ are orthogonal, so the difficulty is with $\bar{\boldsymbol{X}}_j$.

The solution is to give $\bar{\boldsymbol{X}}_j$ and $\tilde{\boldsymbol{X}}_j$ separate effects:

$$\boldsymbol{Y}_j = \bar{\boldsymbol{X}}_j\,\gamma_1 + \tilde{\boldsymbol{X}}_j\,\gamma_2 + \mathbf{1}_j\boldsymbol{U}_j + \boldsymbol{R}_j\ .$$

Now γ_2 has the role of the old γ:
'the estimation is done using only within-group information'.

Often, there are substantive interpretations of the difference between the *within-group effects* γ_2 and the *between-group effects* γ_1.

2. Check random effects of level-1 variables.

See Chapter 5.

4. Check heteroscedasticity.

See Chapter 8.

3,4. Level-1 residual analysis

5,6. Level-2 residual analysis

For residuals in multilevel models, more information is in Chapter 3 of *Handbook of Multilevel Analysis* (eds. De Leeuw and Meijer, Springer 2008) (preprint at course website).

Level-one residuals

OLS within-group residuals can be written as

$$\hat{\boldsymbol{R}}_j = \left(\boldsymbol{I}_{n_j} - \boldsymbol{P}_j\right) \boldsymbol{Y}_j$$

where we define design matrices $\check{\boldsymbol{X}}_j$ comprising $\boldsymbol{X}_j$ as well as $\boldsymbol{Z}_j$ (to the extent that $\boldsymbol{Z}_j$ is not already included in $\boldsymbol{X}_j$) and

$$\boldsymbol{P}_j = \check{\boldsymbol{X}}_j(\check{\boldsymbol{X}}_j'\check{\boldsymbol{X}}_j)^{-1}\check{\boldsymbol{X}}_j' \ .$$

Model definition implies

$$\hat{\boldsymbol{R}}_j = \left(\boldsymbol{I}_{n_j} - \boldsymbol{P}_j\right) \boldsymbol{R}_j \quad :$$

these level-1 residuals are not confounded by $\boldsymbol{U}_j$.

Use of level-1 residuals :

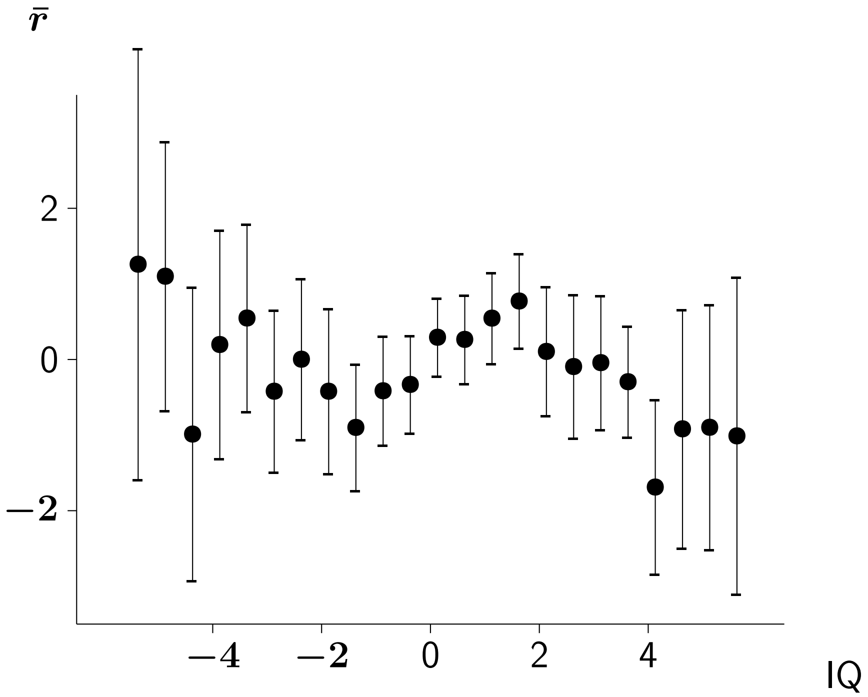
Test the fixed part of the level-1 model using OLS level-1 residuals, calculated per group separately.

Test the random part of the level-1 model using squared standardized OLS residuals.

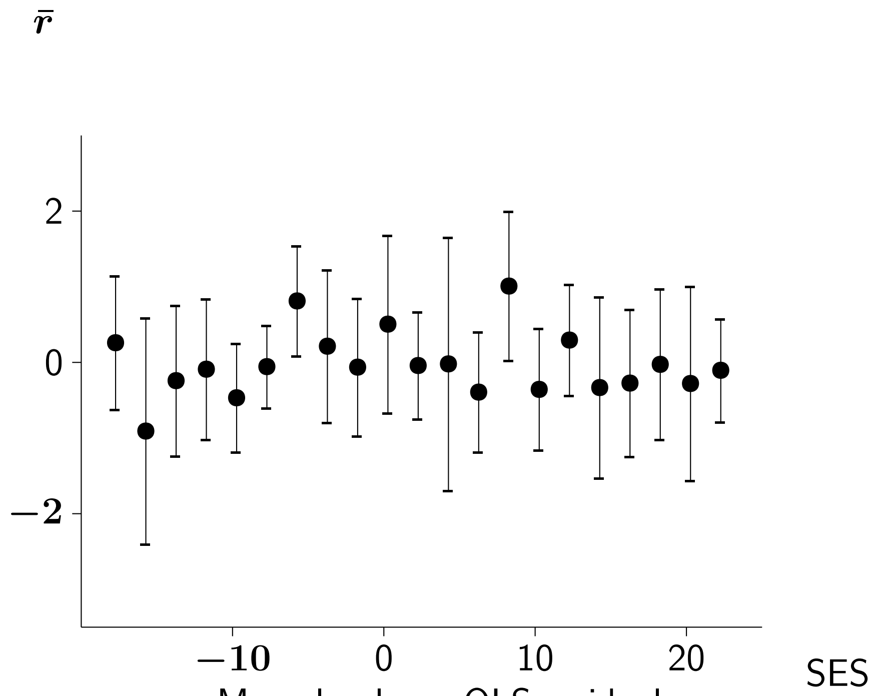
In other words, the level-1 specification can be studied by disaggregation to the within-group level (comparable to a "fixed effects analysis").

The examples of Chapter 8 are taken up again.

Example: model with effects of IQ, SES, sex.



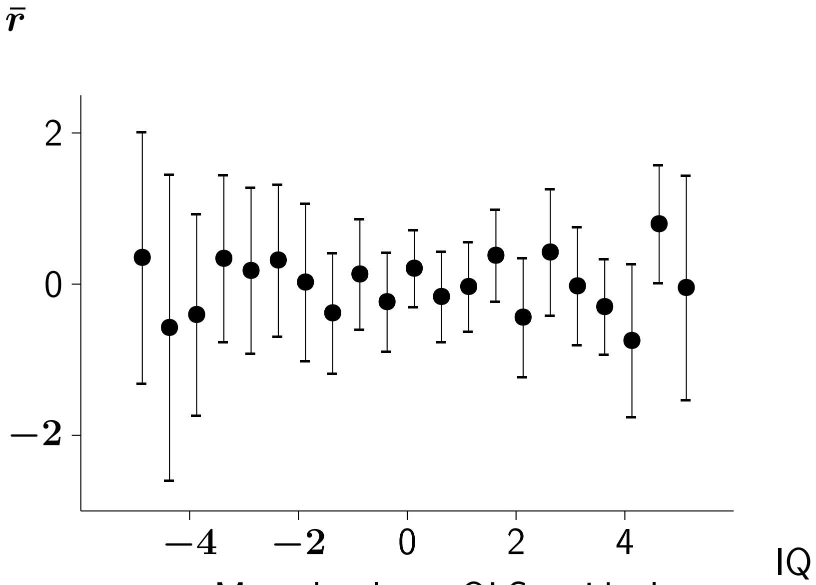
Mean level-one OLS residuals
(bars ∼ twice standard error of the mean)
as function of IQ.



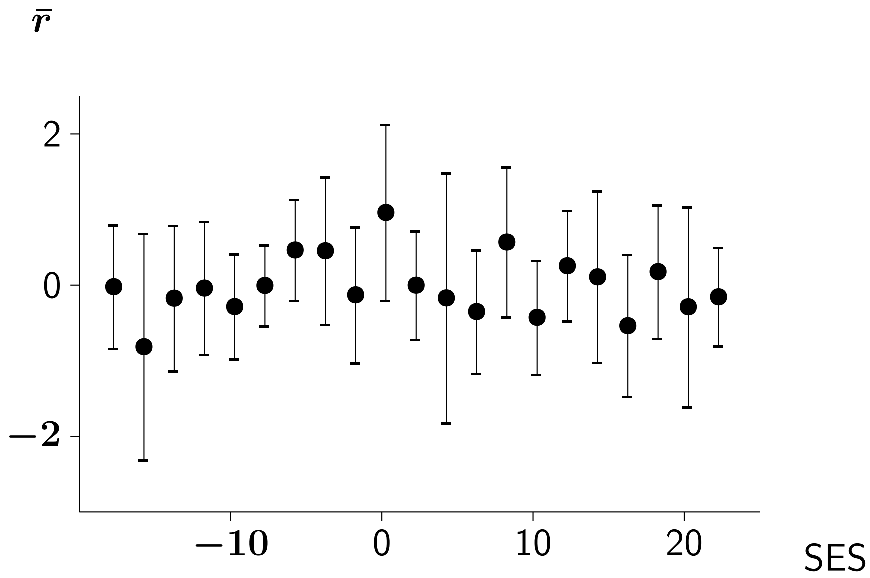
Mean level-one OLS residuals
as function of SES.

This suggest a curvilinear effect of IQ.

Model with effects also of IQ_-^2 and IQ_+^2 .



Mean level-one OLS residuals as function of IQ.



Mean level-one OLS residuals as function of SES.

This looks pretty random.

Are the within-group residuals normally distributed?

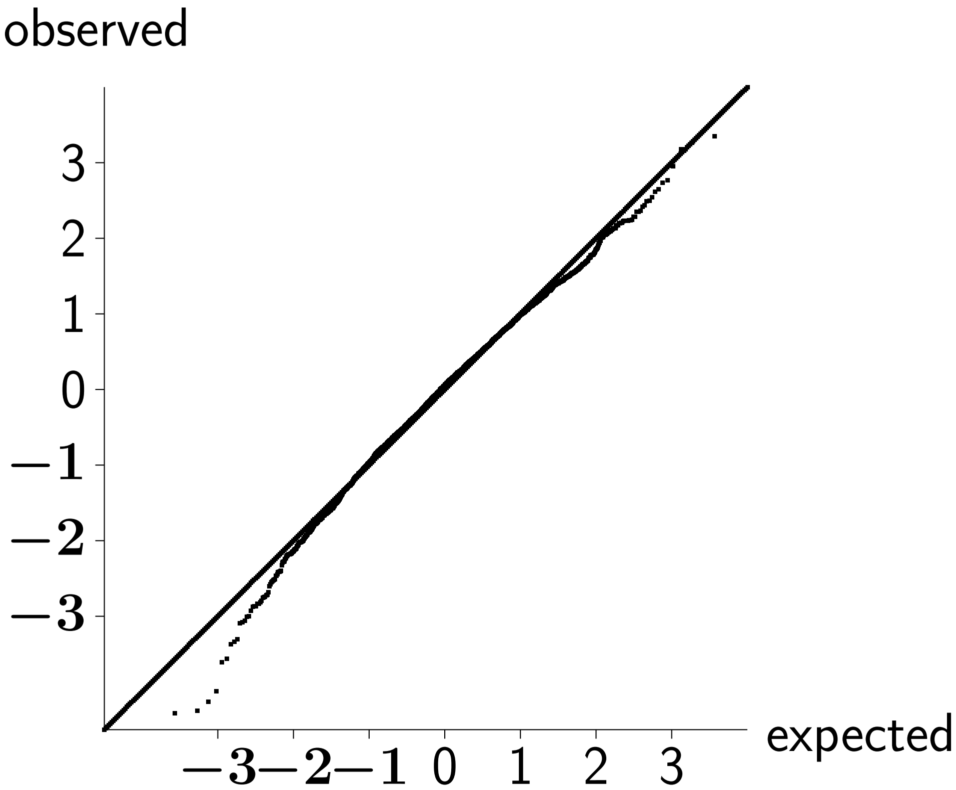


Figure 10.3 Normal probability plot of standardized level-one OLS residuals.

Left tail is a bit heavy, but this is not serious.

Level-two residuals

Empirical Bayes (EB) level-two residuals defined as conditional means

$$\hat{\boldsymbol{U}}_j = E\{\boldsymbol{U}_j \mid \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N\}$$

(using parameter estimates $\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\xi}}$)

$$= \hat{\boldsymbol{\Omega}}\, \boldsymbol{Z}_j'\, \hat{\boldsymbol{V}}_j^{-1}\, (\boldsymbol{Y}_j - \boldsymbol{X}_j \hat{\gamma}_j) = \hat{\boldsymbol{\Omega}}\, \boldsymbol{Z}_j'\, \hat{\boldsymbol{V}}_j^{-1}\, (\boldsymbol{Z}_j\, \boldsymbol{U}_j \;+\; \boldsymbol{R}_j - \boldsymbol{X}_j(\hat{\gamma} - \gamma))$$

where

$$\boldsymbol{V}_j = \mathit{Cov}\, \boldsymbol{Y}_j = \boldsymbol{Z}_j \boldsymbol{\Omega} \boldsymbol{Z}_j' \;+\; \boldsymbol{\Sigma}_j\;,\; \hat{\boldsymbol{V}}_j = \boldsymbol{Z}_j \hat{\boldsymbol{\Omega}} \boldsymbol{Z}_j' \;+\; \hat{\boldsymbol{\Sigma}}_j\;,$$

with $\hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}(\hat{\boldsymbol{\xi}})$ and $\hat{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}_j(\hat{\boldsymbol{\theta}})$.

You don't need to worry about the formulae.

'Diagnostic variances', used for assessing distributional properties of $\boldsymbol{U_j}$:

$$\textit{Cov}\,\hat{\boldsymbol{U}}_j \approx \boldsymbol{\Omega Z}_j'\boldsymbol{V}_j^{-1}\boldsymbol{Z}_j\boldsymbol{\Omega}\ ,$$

'Comparative variances', used for comparing 'true values' $\boldsymbol{U_j}$ of groups:

$$\textit{Cov}\ \left(\hat{\boldsymbol{U}}_j - \boldsymbol{U}_j\right) \approx \boldsymbol{\Omega} - \boldsymbol{\Omega Z}_j'\boldsymbol{V}_j^{-1}\boldsymbol{Z}_j\boldsymbol{\Omega}\ .$$

Note that

$$\textit{Cov}\,(\boldsymbol{U}_j) = \textit{Cov}\,(\boldsymbol{U}_j - \hat{\boldsymbol{U}}_j)\ +\ \textit{Cov}\,(\hat{\boldsymbol{U}}_j)\ .$$

Standardization (by diagnostic variances) :

$\sqrt{\hat{\boldsymbol{U}}_j'\{\widehat{\textit{Cov}}\,(\hat{\boldsymbol{U}}_j)\}^{-1}\hat{\boldsymbol{U}}_j}$ (with the sign reinstated)
is the standardized EB residual.

However,

$$\hat{\boldsymbol{U}}_j'\{\widehat{Cov}\,(\hat{\boldsymbol{U}}_j)\}^{-1}\hat{\boldsymbol{U}}_j \approx \hat{\boldsymbol{U}}_j^{(\mathrm{OLS})\prime}\left(\hat{\boldsymbol{\sigma}}^2(\boldsymbol{Z}_j'\boldsymbol{Z}_j)^{-1} + \hat{\boldsymbol{\Omega}}\right)^{-1}\hat{\boldsymbol{U}}_j^{(\mathrm{OLS})}$$

where $\hat{\boldsymbol{U}}_j^{(\mathrm{OLS})} = (\boldsymbol{Z}_j'\boldsymbol{Z}_j)^{-1}\boldsymbol{Z}_j'\,(\boldsymbol{Y}_j - \boldsymbol{X}_j\hat{\gamma}_j)$

is the OLS estimate of $\boldsymbol{U}_j$, estimated from level-1 residuals $\boldsymbol{Y}_j - \boldsymbol{X}_j\hat{\gamma}_j$.

This shows that standardization by diagnostic variances
takes away the difference between OLS and EB residuals.
Therefore, in checking standardized level-two residuals,
the distinctoin between OLS and EB residuals loses its meaning.

Test the fixed part of the level-2 model using non-standardized EB residuals.

Test the random part of the level-2 model
using squared EB residuals standardized by *diagnostic* variance.

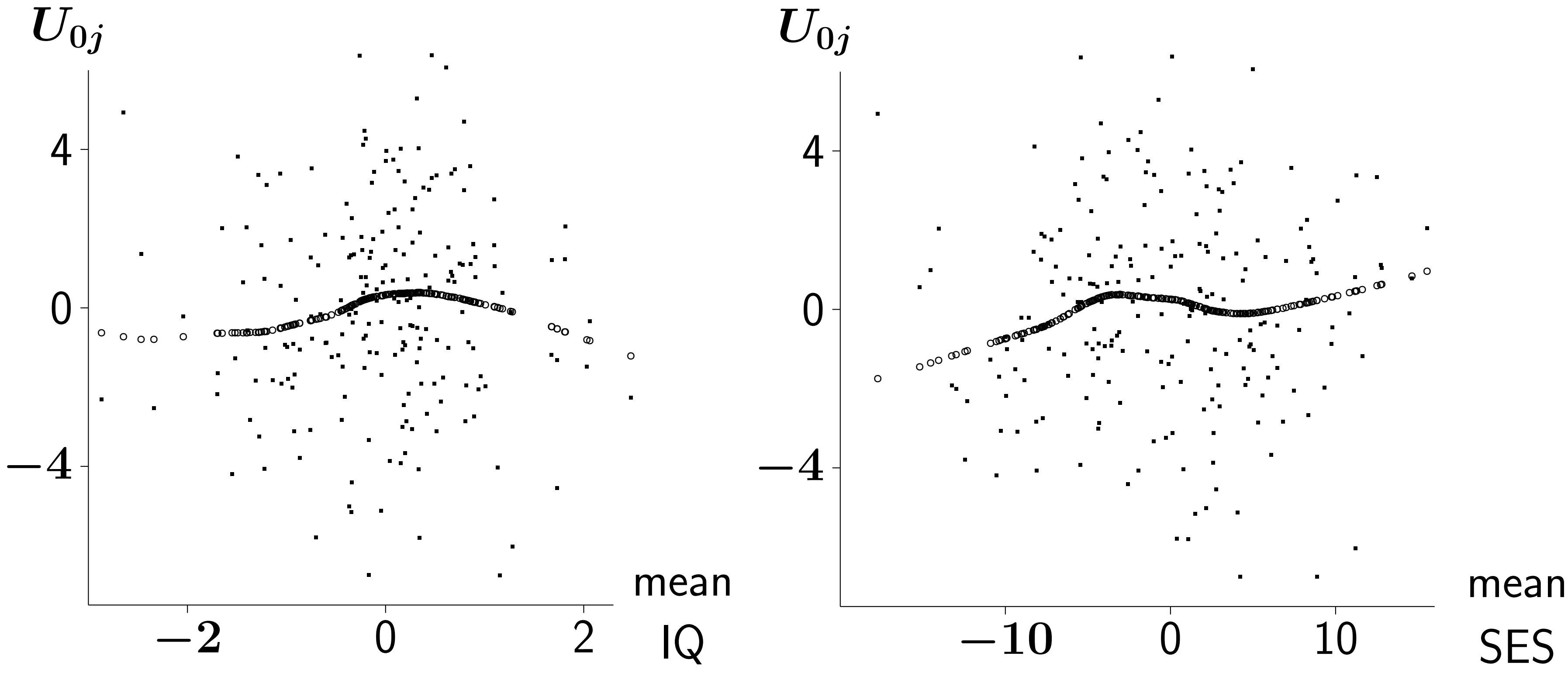


Figure 10.4 Posterior intercepts as function of (left) average IQ and (right) average SES per school. Smooth lowess approximations are indicated by ◦.

The slight deviations do not lead to concerns.

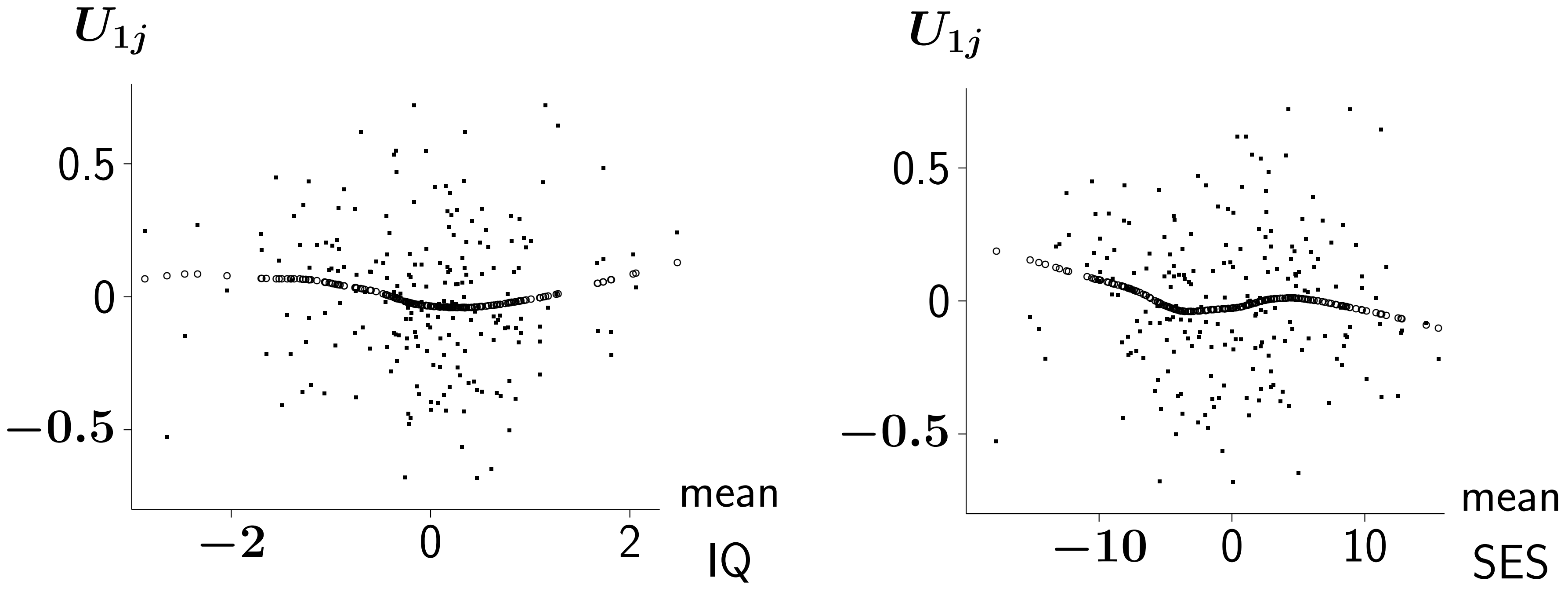


Figure 10.5 Posterior IQ slopes as function of (left) average IQ and (right) average SES per school. Smooth lowess approximations are indicated by ◦.

Again, the slight deviations do not lead to concerns.

Multivariate residuals

The multivariate residual is defined, for level-two unit $\boldsymbol{j}$, as

$$\boldsymbol{Y}_j - \boldsymbol{X}_j \hat{\gamma}.$$

The standardized multivariate residual is defined as

$$\boldsymbol{M}_j^2 = (\boldsymbol{Y}_j - \boldsymbol{X}_j\,\hat{\gamma}_j)'\,\hat{\boldsymbol{V}}_j^{-1}\,(\boldsymbol{Y}_j - \boldsymbol{X}_j\,\hat{\gamma}_j)\,.$$

If all variables with fixed effects also have random effects, then

$$\boldsymbol{M}_j^2 = (\boldsymbol{n}_j - \boldsymbol{t}_j)\,\boldsymbol{s}_j^2 \,+\, \hat{\boldsymbol{U}}_j'\,\{\widehat{\mathit{Cov}}\,(\hat{\boldsymbol{U}}_j)\}^{-1}\,\hat{\boldsymbol{U}}_j\ ,$$

where

$$\boldsymbol{s}_j^2 = \frac{1}{\boldsymbol{n}_j - \boldsymbol{t}_j}\,\hat{\boldsymbol{R}}_j'\,\hat{\boldsymbol{R}}_j\ ,\ \boldsymbol{t}_j = \mathrm{rank}(\boldsymbol{X}_j)\ .$$

This indicates how well the model fits to group $\boldsymbol{j}$.
Note the confounding with level-1 residuals.

If an ill-fitting group does not have a strong effect on the parameter estimates, then it is not so serious.

Deletion residuals

The deletion standardized multivariate residual can be used to assess the fit of group j, but takes out the effect of this group on the parameter estimates:

$$M^2_{(\text{-}j)} = \big(Y_j - X_j\,\hat{\gamma}_{(\text{-}j)}\big)'\,\hat{V}^{-1}_{(\text{-}j)}\,\big(Y_j - X_j\,\hat{\gamma}_{(\text{-}j)}\big)$$

where

$$\hat{V}_{(\text{-}j)} = Z_j\,\hat{\Omega}_{(\text{-}j)}\,Z'_j + \hat{\Sigma}_{(\text{-}j)}\ ,$$

$_{(\text{-}j)}$ meaning that group j is deleted from the data for estimating this parameter.

Full computation of deletion estimates may be computing-intensive,
which is unattractive for diagnostic checks.

Approximations have been proposed:
Lesaffre & Verbeke: Taylor series; Snijders & Bosker: one-step estimates.

The approximate distribution of multivariate residuals, if the model fits well and sample sizes are large, is χ^2, d.f. $= n_j$.

Influence diagnostics of higher-level units

The *influence* of the groups can be assessed by statistics analogous to Cook's distance:
how large is the influence of this group on the parameter estimates?

Standardized measures of influence of unit j on fixed parameter estimates :

$$C_j^F = \frac{1}{r} \left(\hat{\gamma} - \hat{\gamma}_{(\text{-}j)}\right)' \hat{S}_{F(\text{-}j)}^{-1} \left(\hat{\gamma} - \hat{\gamma}_{(\text{-}j)}\right)$$

where S_F is covariance matrix of fixed parameter estimates, and $_{(\text{-}j)}$ means that group j is deleted from the data for estimating this parameter.

on random part parameters :

$$C_j^R = \frac{1}{p} \left(\hat{\eta} - \hat{\eta}_{(\text{-}j)}\right)' \hat{S}_{R(\text{-}j)}^{-1} \left(\hat{\eta} - \hat{\eta}_{(\text{-}j)}\right) \ ,$$

combined :

$$C_j = \frac{1}{r + p} \left(r C_j^F + p C_j^R\right) .$$

Values of C_j larger than 1 indicate strong outliers.
Values larger than $4/N$ may merit inspection.

Table 10.1 the 20 largest influence statistics, and p-values for multivariate residuals,
of the 211 schools; Model 4 of Chapter 8 but without heteroscedasticity.

School	n_j	C_j	p_j
182	9	0.053	0.293
107	17	0.032	0.014
229	9	0.028	0.115
14	21	0.027	0.272
218	24	0.026	0.774
52	21	0.025	0.024
213	19	0.025	0.194
170	27	0.021	0.194
67	26	0.017	0.139
18	24	0.016	0.003

School	n_j	C_j	p_j
117	27	0.014	0.987
153	22	0.013	0.845
187	26	0.013	0.022
230	21	0.012	0.363
15	8	0.012	0.00018
256	10	0.012	0.299
122	23	0.012	0.005
50	24	0.011	0.313
101	23	0.011	0.082
214	21	0.011	0.546

School 15 does not survive Bonferroni correction: 211 $\times$ 0.00018 = 0.038. Therefore now add the heteroscedasticity of Model 4 in Chapter 8.

Table 10.2 the 20 largest influence statistics, and p-values for multivariate residuals,
of the 211 schools; Model 4 of Chapter 8 with heteroscedasticity.

School	n_j	C_j	p_j
213	19	0.094	0.010
182	9	0.049	0.352
107	17	0.041	0.006
187	26	0.035	0.009
52	21	0.028	0.028
218	24	0.025	0.523
14	21	0.024	0.147
229	9	0.016	0.175
67	26	0.016	0.141
122	23	0.016	0.004

School	n_j	C_j	p_j
18	24	0.015	0.003
230	21	0.015	0.391
169	30	0.014	0.390
170	27	0.013	0.289
144	16	0.013	0.046
117	27	0.013	0.988
40	25	0.012	0.040
153	22	0.012	0.788
15	8	0.011	0.00049
202	14	0.010	0.511

School 15 now does survive the Bonferroni correction: $211 \times 0.00049 = 0.103$.
Therefore now add the heteroscedasticity of Model 4 in Chapter 8.
Another school (108) does have poor fit $\boldsymbol{p = 0.00008}$, but small influence ($\boldsymbol{C_j = 0.008}$).

Leaving out ill-fitting schools does not lead to appreciable differences in results.

The book gives further details.