Multilevel Modelling

University of Zürich
PD Dr. Conrad Ziller
conrad.ziller@uni-due.de
https://conradziller.com

28.10.2021

Who am I?

- MA Political Science at U Tübingen
- PhD and Habilitation at U Cologne
- Since 10/2020 Akademischer Rat (eq. Lecturer) U Duisburg-Essen

- Research interests:
  - Migration and integration
  - Policy effects
  - Political and social trust
  - Quantitative methods

Who are you?

- 5 min breakout session in pairs, talk about the following three questions, then introduce your breakour session partner in the audiance:
  - What is your name?
  - Where did you grow up? With which city/area do you identify most?
  - What is an interesting/outstanding characteristic about yourself?

Schedule for the course

DAY 1

- Introduction, Fundamentals
- Random Intercept and Random Slope models
- Assumptions and diagnostics
- Cross-level interactions
- Centering

DAY 2

- Fixed Effects models
- 3-Level models (and more)
- Cross-classified models
- Logit and other link functions

DAY 3

- Growth curve models and multilevel SEM
- Merits and pitfalls of multilevel models
- Discussion of your projects

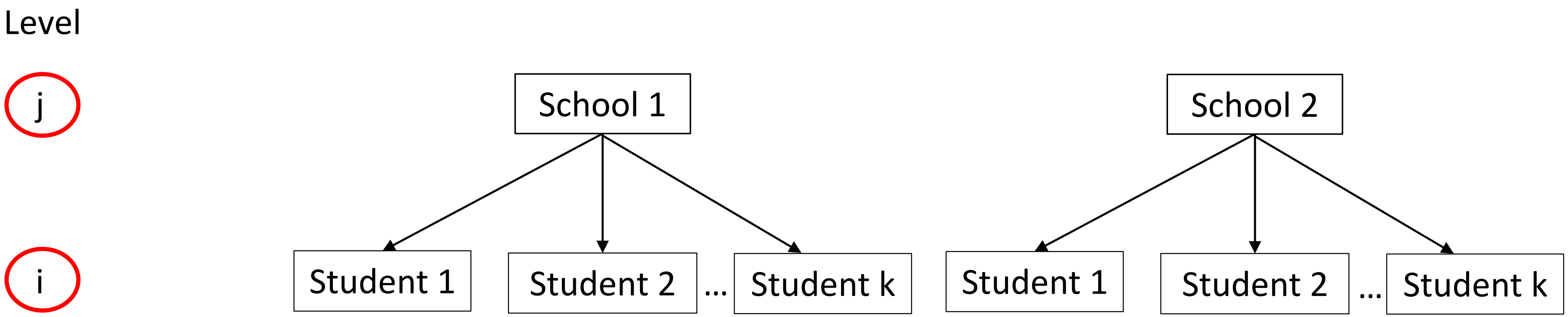
Schedule for Day 1

- **Multilevel data and handling non-independent observations**
- Fundaments of multilevel modelling
- Random Intercept and Random Slope models
- Assumptions and diagnostics
- Cross-level interactions
- Centering

Multilevel data

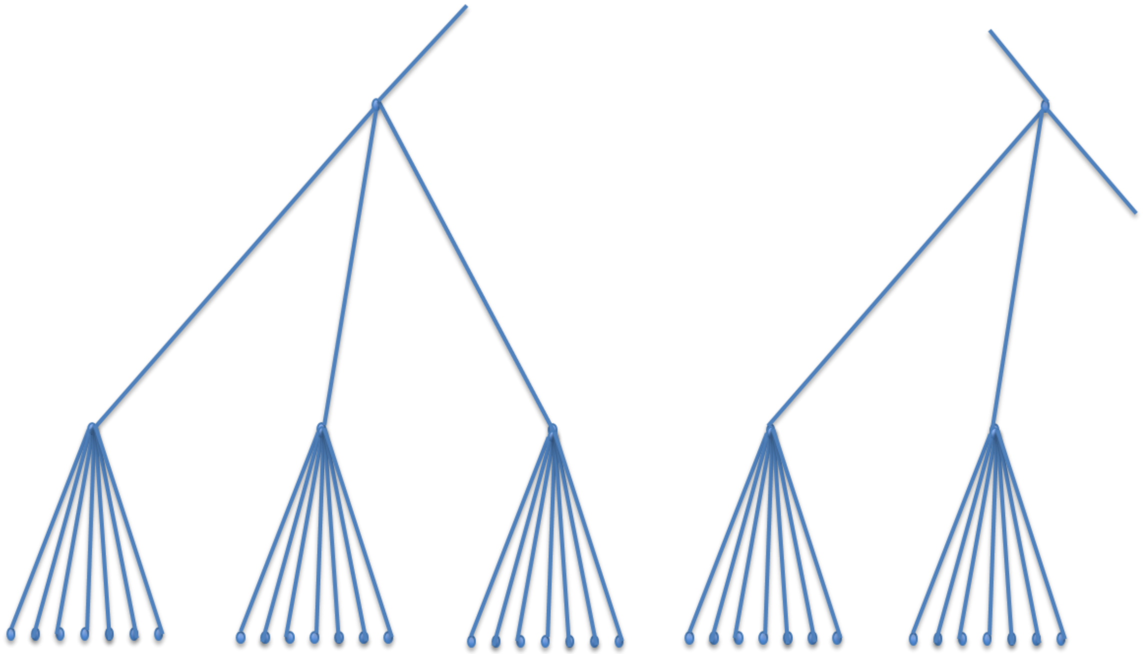
Data is hierarchical / grouped / clustered:

- Temporarily structured: data measured over time
- Spatially structured data, e.g. students in schools
- Within individual variation, e.g. data from factorial survey experiments

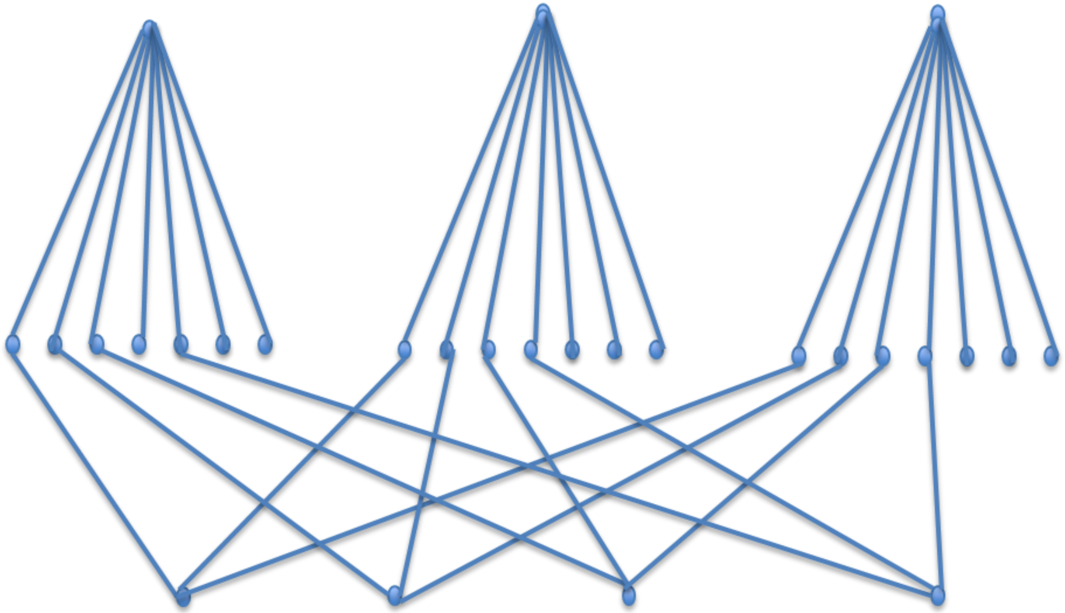


Multilevel data

Grouped data



Complex grouping



Multilevel data

Data structure

ID_i	ID_j	x1	x2	x3	z1	z2	y
1	1	1	3	1	2	4	2
2	1	2	2	2	2	4	5
3	2	3	1	4	3	1	1
4	2	2	2	1	3	1	2
5	2	1	4	3	3	1	3
6	3	5	4	1	4	2	1

Multilevel data

- Violation of uncorrelated errors → serial correlation
  - Indicated by intra-class correlation coefficient (ICC): The larger the ICC the greater the similarities within the groups (and the larger the average differences between groups)
  - Breusch Pagan / Wooldridge test

- Serial correlation: coefficients might be unbiased, but the standard errors are in any case biased
- Affects statistical testing procedures: underestimation of the standard errors
- Leads to detecting statistically significant relationships that actually do not exist

Multilevel data

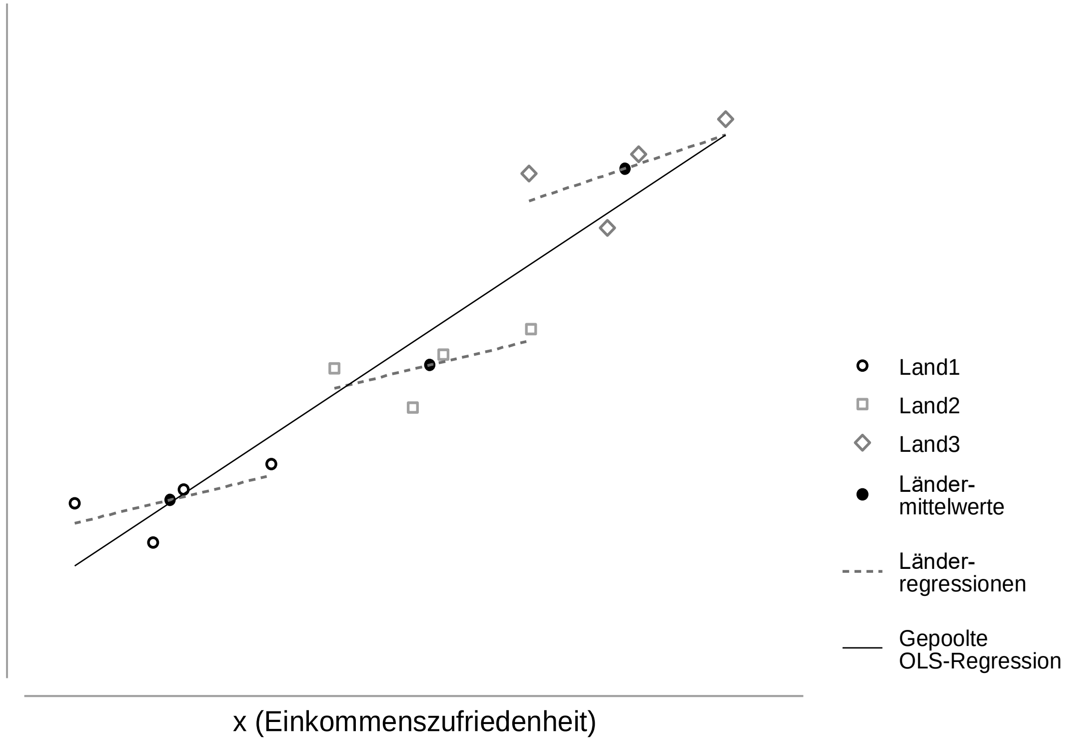
- Violation of the exogeneity assumption (error term e and predictor x are uncorrelated: $E(e|x) = 0$)

→ If group membership carries relevant unobserved characteristics, then biased regression coefficients!

- Relevant means correlated with outcome variable and predictor variable of interest

Multilevel data

- Example: Group membership (country) correlated with satisfaction with government (y) AND satisfaction with income (x)
- Group differences represent unobserved country characteristics (e.g. unemployment rates) that are „transported“ by the individual-level variable



Remedies addressing…

Serial correlation

- Control for factors that causally underlie the process of serial correlation
- Estimation of cluster - / panel-robust standard errors
- Fixed-Effect Models
- Random-Effects / Multilevel Models

Unoverserved heterogeneity

- Fixed-Effect Models
- Random-Effects / Multilevel Models with group-level variables

Schedule for Day 1

- Multilevel data and handling non-independent observations
- **Fundaments of multilevel modelling**
- Random Intercept and Random Slope models
- Assumptions and diagnostics
- Cross-level interactions
- Centering

Fundamentals – When to do multilevel modeling?

Statistical considerations:

- Standard methods such as OLS ignore the hierarchical structure of the data

Substantial Interest:

- Interest in context/group effects:
  - Interest in effect of group variables
  - Interest in the variance part
  - Interest in interaction between group and individual-level variables

Fundamentals – Levels versus variables

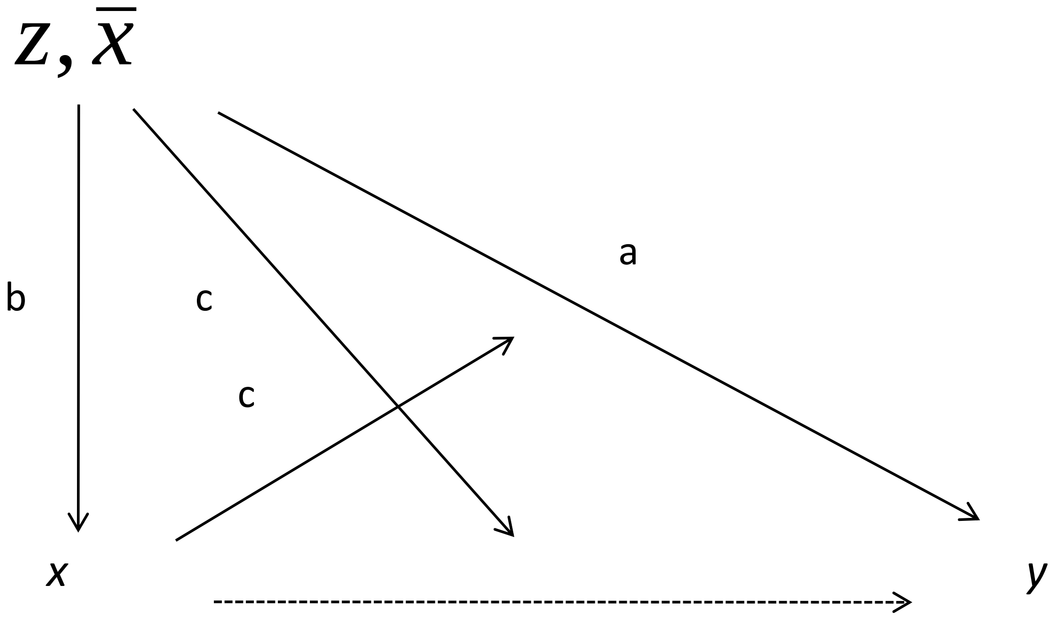
- Various labels: Hierarchical Linear Models, Random Coefficient Models, Mixed Models

- Level criteria:
  - Vast number of categories
  - Without intrinsic information content
  - Random selection from a "case universe" (interchangeability)

- Level vs. Variables
  - Example 1: Organisations in which person works (employees in organisations)
  - Example 2: Social class (people in social classes)
  - Example 3: Neighborhoods
  - Example 4: East Germany vs. West Germany

Fundamentals

Different contextual effects



Z= Global variable

$\bar{x}$ = Aggregated Variable

a Direct contextual effects

b Perception-mediated effects

c Cross-level interactions

Group session #1

- You will be assigned to breakout rooms in pairs and draw
- Use the whiteboard function in Zoom or https://webwhiteboard.com/
- Draw the multilevel relationships you are interested in (e.g. own project)
- Save a screenshot or the web whiteboard link to share your picture with all participants

Schedule for Day 1

- Multilevel data and handling non-independent observations
- Fundaments of multilevel modelling
- **Random Intercept and Random Slope models**
- Assumptions and diagnostics
- Cross-level interactions
- Centering

Overview

- Basic ML-models
  - Random Intercept Model
  - Random Slope Model
- Further topics: Model estimation, model assumptions, model fit, binary DV, interactions, mediation.

Overview

Typical procedure

1. (Empty model)
2. Random Intercept Model, individual variables only
3. Random Intercept Model, individual variables and context variables
4. Random Slopes Model
5. Random Slopes Model, with cross-level interactions

- Example: 10000 individuals in 20 countries, DV: social trust, IVs: age, education, income + GDP/capita

Random Intercept Model

Step 1: Regression for country j

$$y_{ij} = \boxed{\beta_{0j}} + \beta_{1j}x_{ij} + e_{ij}$$

Deviation from regression lines

Trust of person i in country j

Effect of income

Average trust in country j

Random Intercept Model

Step 2: Specification of country-specific intercepts

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Overall notation:

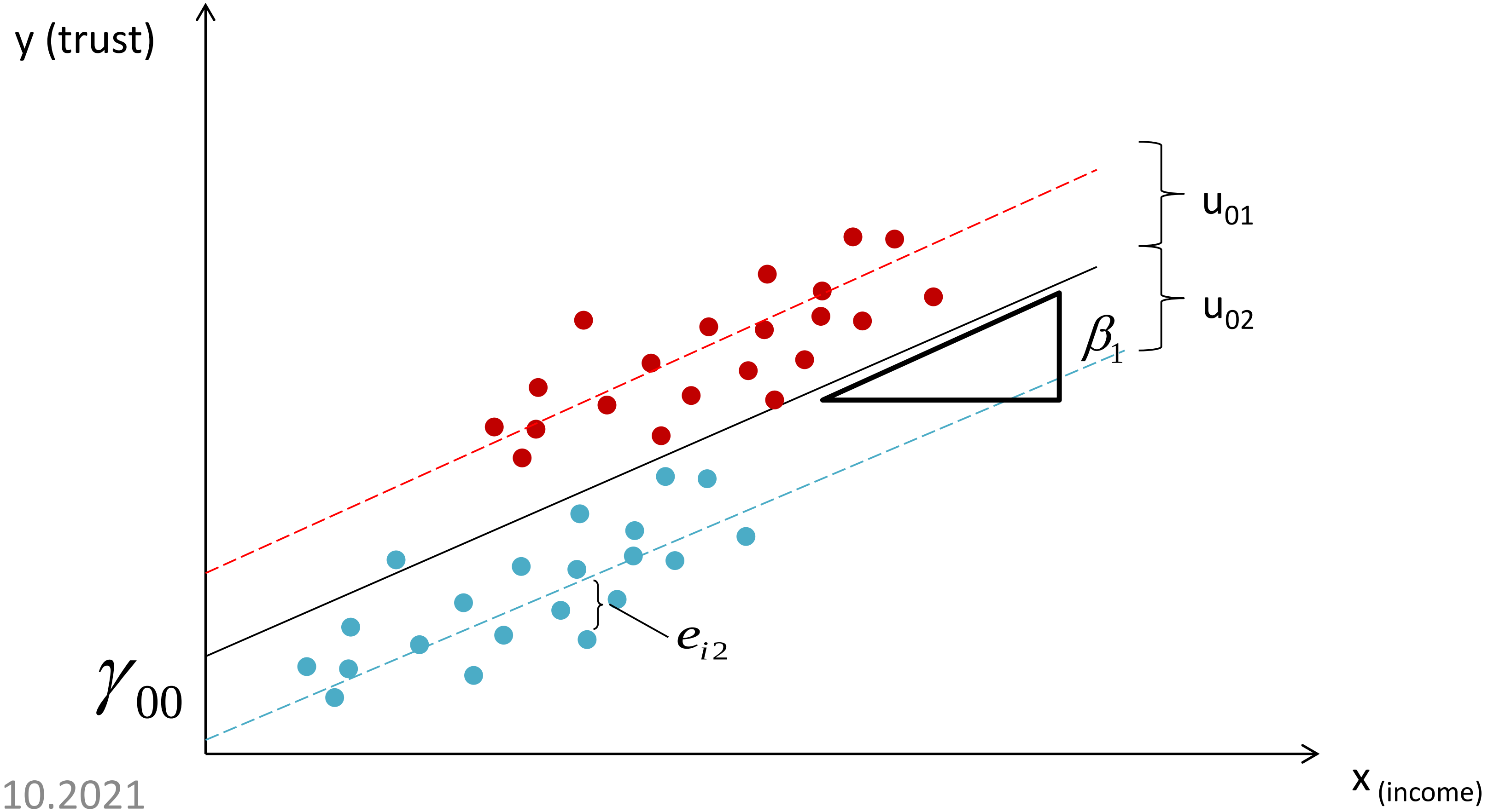
Average trust across countries („Grand Mean“ or „Grand Intercept“)

Deviation of a country‘s trust level from the Grand Mean

$$y_{ij} = \underbrace{\gamma_{00} + \beta_1 x_{ij}}_{\text{Fixed Part / Fixed Effects}} + \underbrace{u_{0j} + e_{ij}}_{\text{Random Part / Random Effects}}$$

Random Intercept Model

$$y_{ij} = \gamma_{00} + \beta_1 x_{ij} + u_{0j} + e_{ij}$$



Random Intercept Model

- Variance can be decomposed:

$$\text{Total variance } = \sigma_{u_0}{}^2 + \sigma_e{}^2$$

- Intra-Class-Correlation (ICC)

$$\rho = \frac{\sigma_{u_0}{}^2}{\sigma_{u_0}{}^2 + \sigma_e{}^2}$$

  – Proportion of variance at the grouping level
  – Mean correlation of y-values when pairs of the same group are considered

Rule of thumb:
$\rho \leq 0{,}05$ small amount of L2-varince
$\rho \cong 0{,}10$ medium amount of L2-varince
$\rho \geq 0{,}15$ large amount of L2-varince

- Always calculate ICC from empty model before ML analysis

Random Intercept Model

- Introduction of macro variables serves to reduce (or explain) the random intercept variance:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\boxed{z_j} + u_{0j}$$

- Summarized notation:

$$y_{ij} = \gamma_{00} + \beta_1 x_{ij} + \gamma_{01} z_j + u_{0j} + e_{ij}$$

e.g. effect of GDP/capita on average level of social trust

Stata

```
mixed DV IV1 IV2 … ||id:
```

R

```
lmer(DV ~ 1 + IV1 + IV2 + (1|id), data = data)
```

Random Slope Model

- Not only the mean value of the DV, but also the effect size of individual IVs can vary across contexts
- Random slope (RS) = group-specific slope of the regression line.
- There are both theoretical (e.g. there are reasons why a relationship is not equally strong in all contexts) and empirical reasons to model RS.
- Model comparisons (likelihood ratio test) provide information on the relevance of RS

Random Slope Model

Step 1: Regression for country j

$$y_{ij} = \boxed{\beta_{0j}} + \beta_{1j}x_{ij} + e_{ij}$$

Deviation from regression lines

Trust of person i in country j

Effect of income

Average trust in country j

Random Slope Model

- Step 2: Specification of country-specific intercepts and **country-specific slopes**

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + u_{0j} \qquad \beta_{1j} = \gamma_{10} + u_{1j}$$

Average income effect across all countries

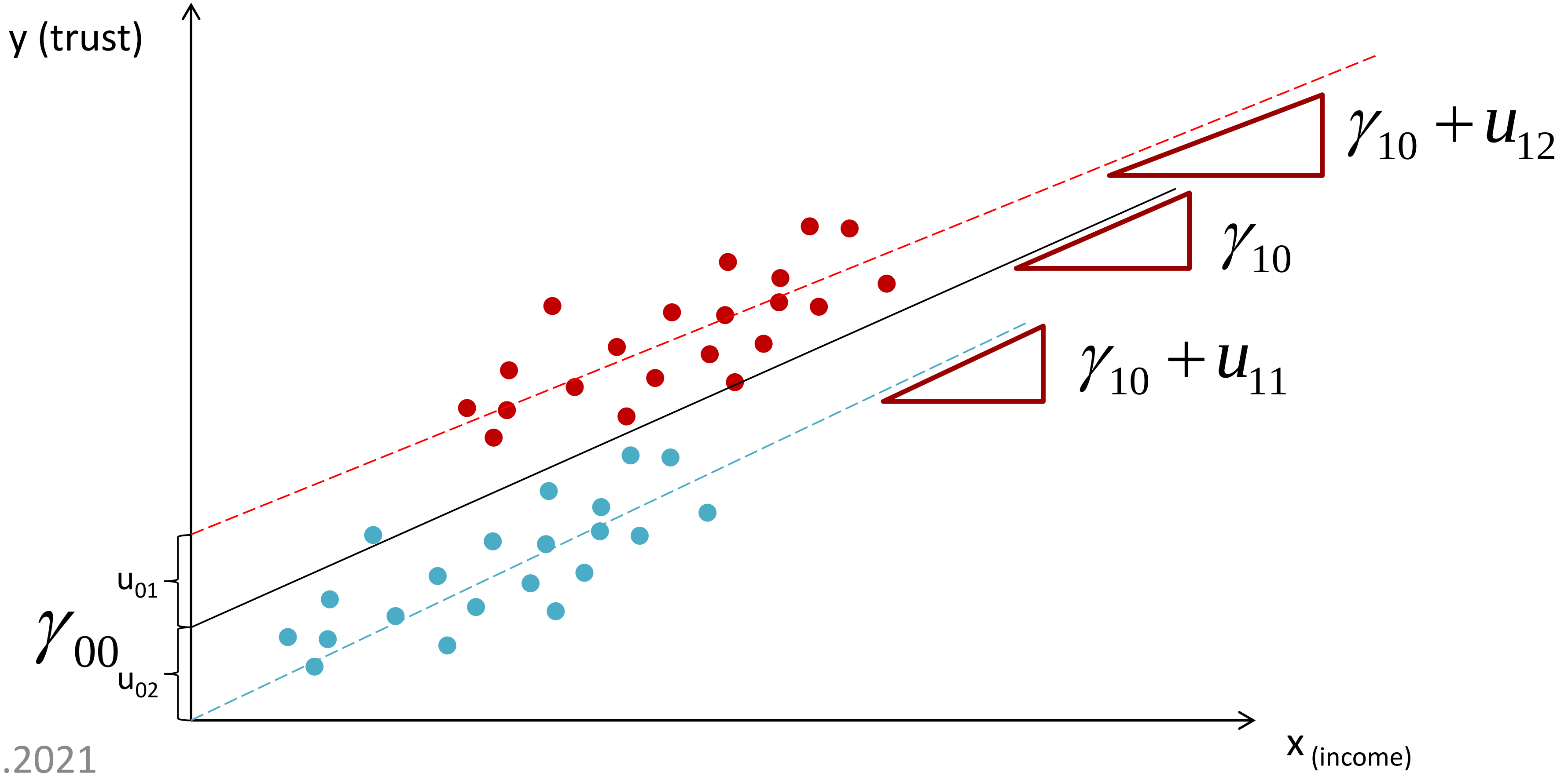
Deviation country j

- Overall notation:

$$y_{ij} = \underbrace{\gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} z_j}_{\text{Fixed Part / Fixed Effects}} + \underbrace{u_{1j} x_{ij} + u_{0j} + e_{ij}}_{\text{Random Part / Random Effects}}$$

Random Slope Model

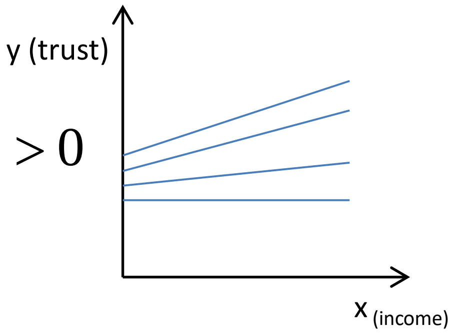
$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}z_j + u_{1j}x_{ij} + u_{0j} + e_{ij}$$



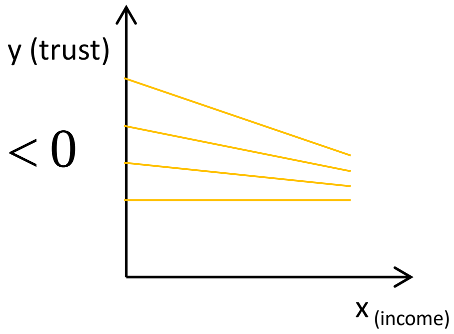
Random Slope Model

- Further specification: variance structure of random effects (correlation between intercept and slopes)

$\sigma_{u_0 u_1} > 0$



$\sigma_{u_0 u_1} < 0$



- Testing the relevance: Likelihood-Ratio-Test

Stata

```
mixed DV IV1 IV2 … ||id: IV1 , cov(un)
```

R

```
lmer(DV ~ 1 + IV1 + IV2 + (1 + IV1|id), data = data)
```

Group session #2

- You will be assigned to breakout rooms in pairs and draw
- Download the data and preliminary syntax files from https://github.com/conrad-ziller/ml-course/tree/main/data
- Group work
  - Decide on which software you use
  - Get an overview of the syntax and talk about the necessary steps to make the syntax work
  - Share your screen and make it work
  - Share the file via chat (a filled out file will be provided at the end of the day)
- ~25 min time

Schedule for Day 1

- Multilevel data and handling non-independent observations
- Fundaments of multilevel modelling
- Random Intercept and Random Slope models
- **Assumptions and diagnostics**
- Cross-level interactions
- Centering

Model Estimation

- Maximum-Likelihood estimation
  - Idea: Find the combination of parameters that take the highest value on the likelihood function, i.e. for which the realization of the observed data is most plausible.

$$\hat{\theta}_{mle} = \underset{\theta \in \Theta}{\arg\max}\, \hat{\ell}(\theta \mid x_1, \ldots, x_n)$$

  - Deviation = difference between model and data
    Iterative procedure:
    - Start from a set of initial values
    - Change parameters that increase the likelihood function
    - Continue until change is maximum = convergence

  - Reasons for non-convergence
    - Misspecified model (overfitting)
    - Poor initial values
    - Parameters too small (close to 0)

Model Estimation

- Nested Models (Model 1 is a subset of Model 2)
  - –Differ e.g. in only one parameter
  - –Comparison via deviance (= -2*LogLikelihood): lower deviance, better model
  - –Likelihood ratio test compares deviance in relation to differences in degrees of freedom
  - –Significant test result shows that the extended model is better

Stata

```
mixed DV IV1 IV2 … ||id:
eststo M1

mixed DV IV1 IV2 … ||id: IV1
eststo M2

lrtest M1 M2
```

R

```
M1 <- lmer(DV ~ 1 + IV1 + IV2 + (1|id) + (0 + IV1|id),
data = data)


M2 <- lmer(DV ~ 1 + IV1 + IV2 + (1 + IV1|id), data =
data)

anova(M2, M1)
```

Model Comparison

- Non-Nested Models
  - AIC (Akaike Information Criterion) = deviance + number of parameters)
  - BIC (Bayes Information Criterion) = deviance + no.par.*ln(number of cases)
  - Consider complexity of models
  - AIC if the number of cases is the same and BIC if it differs
- Proportion of explained variance
  - For each level seperately, in relation to the empty model
  - More difficult to calculate for RS (see Hox 2010: pp.76)

$$R^2_{INDIV} = \frac{\sigma_e^2(empty) - \sigma_e^2}{\sigma_e^2(empty)} \qquad R^2_{CONTEXT} = \frac{\sigma_{u_0}^2(empty) - \sigma_{u_0}^2}{\sigma_{u_0}^2(empty)}$$

Model Assumptions

- *Linearity* of the relationship between x_{ij} and y_{ij}
- Error term and covariates are not correlated: $\mathrm{Cov}(e_{ij}\,, x_{ij}) = 0$
- *Homoscedasticity*: $\mathrm{Var}(e_{ij} \mid x_{ij}) = \sigma^2$
- Error terms are normally distributed with expected value 0:

$$e_{ij} \sim N(0, \sigma_e^2)$$
$$u_{0j} \sim N(0, \sigma_{u_0}^2)$$
$$u_{1j} \sim N(0, \sigma_{u_1}^2)$$

- Correlated error terms within groups *j*: $\mathrm{Corr}(e_{ij}\,, e_{i'j}) = \rho$

Model Assumptions

- Testing of model assumptions via residual diagnostics
- Multiple residuals, seperate diagnostics desirable
- Stepwise approach to diagnostics
- Observations are a common function of the residuals of different levels

Bottom-up strategy:

(1) First diagnose the residuals of the first level

(2) then diagnose the random effects

Stata

Level 1

```
predict yhat
predict res1, rstandard
qnorm res1, name(res1)
twoway scatter res1 yhat, yline(0)
```

Level 2 (with one RS)

```
predict u1 u0, reffects
preserve
collapse yhat u0 u1, by(id_c)
histogram u0, normal name(u0)
histogram u1, normal name(u1)
twoway scatter u0 yhat, yline(0) name(s0) mlabel(id_c)
twoway scatter u1 yhat, yline(0) name(s1) mlabel(id_c)
restore
```

Model Assumptions

- Possible reasons for violated assumptions
  - Non-normally distributed continuous DV or IV
  - Outliers
  - …

- Countermeasures
  - Transformation of variables (e.g. via ladder command in Stata)
  - Interaction terms with heteroskedasticity
  - Remove outliers from model
  - Dummy for outliers

Checklist Modelspecifications

- Is the functional form of the relationship between x and y linear?
- Is the functional form of the relationship between x and y homoscedastic?
  - ➢If not, consider interaction terms
- Relevant control variables in the model?
- Is there unexplained intercept variance at the context level?
  - ➢If yes, is it of theoretical interest?
  - ➢If yes, inclusion of context variables

Checklist Modelspecifications

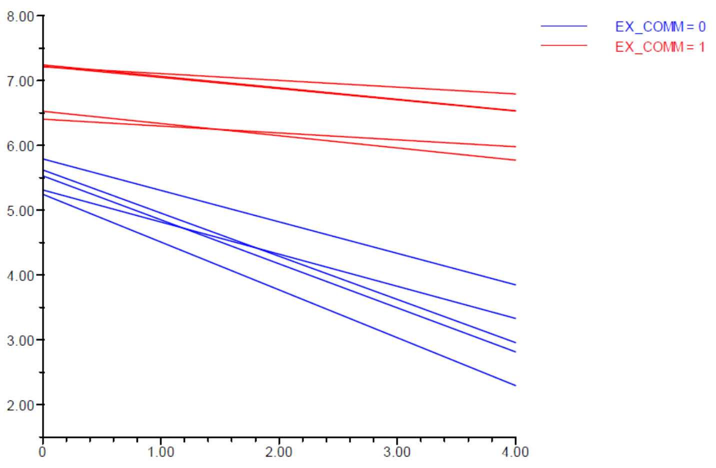
- Does the effect of individual variables vary significantly between contexts?
  - ➢ If yes, random slopes
  - ➢ If necessary, estimate covariances between random effects arbitrarily (mixed option cov(unstructured))
- Should the variance of the random slopes be explained?
  - ➢ If yes, inclusion of multilevel interactions

Schedule for Day 1

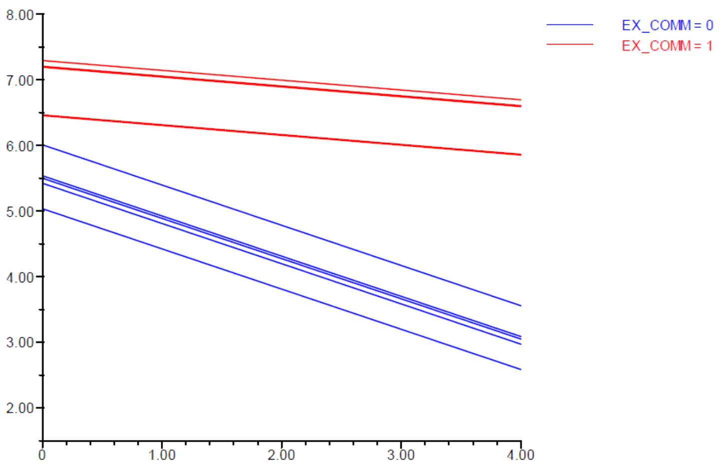
- Multilevel data and handling non-independent observations
- Fundaments of multilevel modelling
- Random Intercept and Random Slope models
- Assumptions and diagnostics
- **Cross-level interactions**
- Centering

Cross-Level-Interactions

- Interaction between context and individual variable
- Idea: Context feature explains different effect of individual feature across contexts (random slope)
- A priori (usually) identify significant random slope



Interaction with RS



Interaction without RS

Cross-Level-Interactions

- Specification of country-specific intercepts and slopes

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

- With interaction:

$$\beta_{1j} = \gamma_{10} + \boxed{\gamma_{11} z_j} + u_{1j}$$

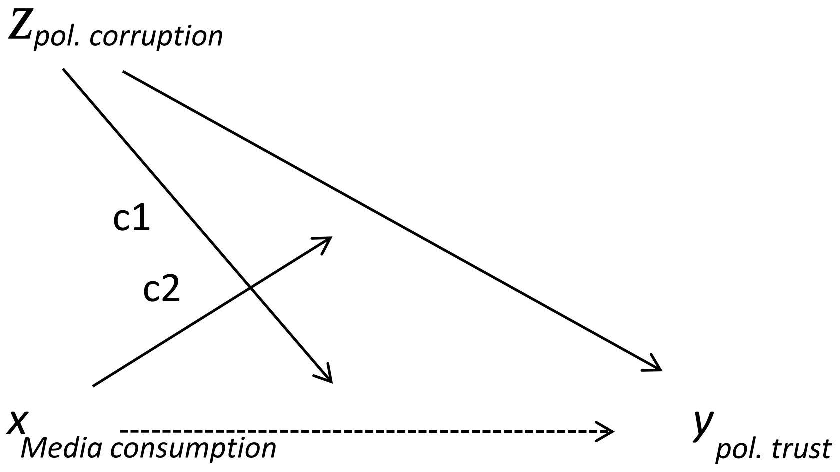
- Overall notation

$$y_{ij} = \gamma_{00} + \underbrace{\gamma_{10} x_{ij} + \gamma_{01} z_j + \gamma_{11} x_{ij} z_j}_{\text{Fixed Part / Fixed Effects}} + \underbrace{u_{1j} x_{ij} + u_{0j} + e_{ij}}_{\text{Random Part / Random Effects}}$$

Cross-Level-Interactions

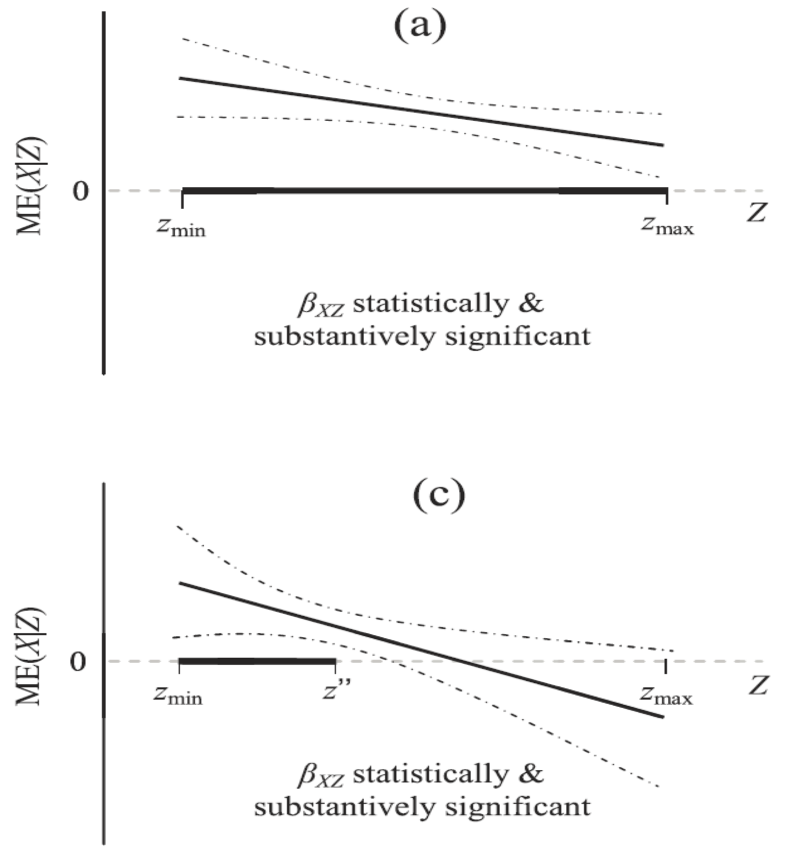
- Hypotheses should be symmetrical (if theoretically reasonable)
  - H1: Political corruption reduces political trust, especially when media consumption is high. (c2)
  - H2: Media consumption reduces political trust, especially in contexts characterized by high political corruption. (c1))

- Statement on high and low values of the moderator often useful
  - H2a: In contexts with ***high*** political corruption, the relationship between media consumption and political trust is ***negative***.
  - H2b: In contexts with ***low*** political corruption, the relationship between media consumption and political trust is ***positive***.



Cross-Level-Interactions

- Always estimate both interaction term and main effects in the model
- Symmetrical interpretation of the results along the hypotheses formulated.
- Is the correlation significant for all values of the moderator? If not, what does this mean for the hypotheses?



Berry et al. 2012: 661

Stata

```
//Form interaction term beforehand
gen x_z=x*z
mixed y x z x_z ||id:


//better interact within the model per # (simplifies post-
estimation)
mixed y x z c.x#c.z ||id:

margins, dydx(x) at(z = (0 (0.1) 1))
marginsplot,  yline(0)
margins, dydx(z) at(x = (0 (0.1) 1))
marginsplot,  yline(0)
```

Schedule for Day 1

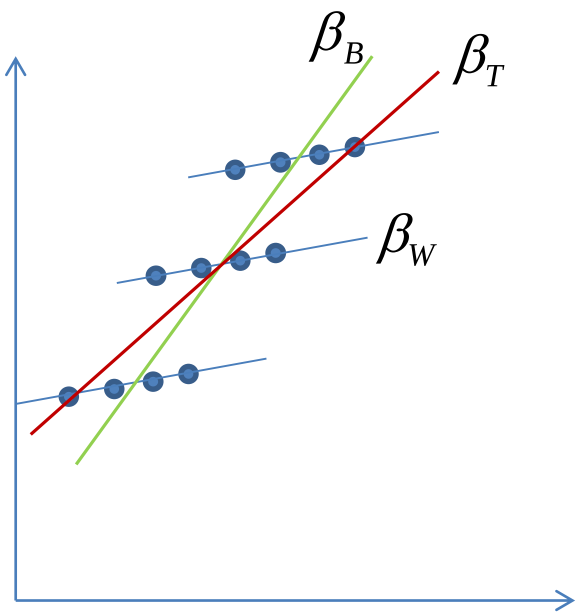
- Multilevel data and handling non-independent observations
- Fundaments of multilevel modelling
- Random Intercept and Random Slope models
- Assumptions and diagnostics
- Cross-level interactions
- **Centering**

Centering

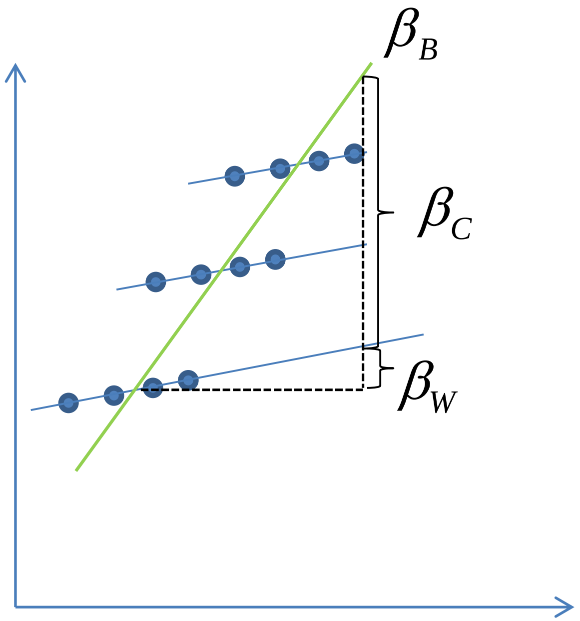
- Version 1: At the overall mean or grand mean (also CGM)
  - Continuous variables receive a meaningful zero point
  - Constant can be interpreted as an estimated value of the AV for persons who have a mean expression on all characteristics.
  - possibly reduces multicollinearity problems
  - all other parameter estimates remain identical (as in the uncentred case)

- Version 2: At the group mean (also CWC)
  - only for individual variables
  - Consequence: Between-group variance is removed from variable
  - Constant is an estimate of the AV for individuals who have an alignment with the group mean on all characteristics
  - Changes correlation structure of data; coefficient estimates especially of level-2 variables change

Context Effect of Group Mean



blue = Within-Effect
green = Between-Effect
red = Total Effect (mixed)



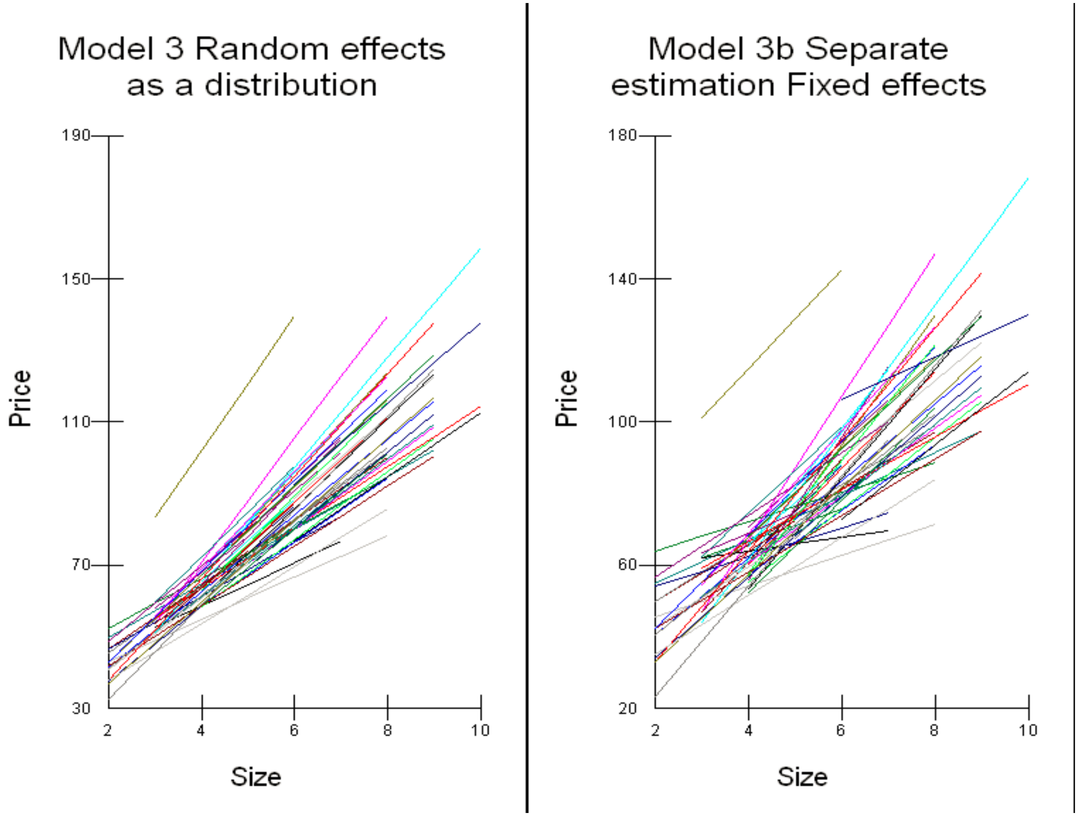
Between-Effect = Within-Effect
+ context effect

Centering and Effect Interpretation

- Centering at the Grand Mean (CGM)
  - –Effect of x = β(within)
  - –Effect of x(line) the group mean = β(context)
  - –β(between) = β(within) + β(context)
- Centering at the Group Mean (CWC)
  - –Effect of x = β(within) *[without „shrinkage"]*
  - –Effect of x(line) the group mean= β(between)
  - –β(context) = β(between) - β(within)

Shrinkage

- Regression lines across individual contexts "shrink" towards the Grand Mean
- The fewer the observations at level 1, the stronger the shrinkage
- Certainty of estimation is to some extent borrowed from large clusters



from: Jones & Subramanian 2009, MlWin Traning Manual

Centering

- When to center on the overall mean?
  - Mostly useful, as it facilitates interpretation of the concept and context effects.
  - If substantial interest in level 2 variable (or interaction at level 2), as it co-controls for composition effects of level 1.
- When to center on the group mean?
  - Content reasons: In the case of poorly comparable group means (e.g. centre income at the country mean).
  - If there is substantial interest in level 1 variables, as there is no shrinkage.
  - In the case of cross-level interaction, as there is no shrinkage at level 1 and level 2 variables here are between and not a mixture of between and within.
  - For level-1 interactions

Stata

```
ssc install center

//grand-mean
center UV1 UV2 , pre(cgm_) mean(mgm_)

//group-mean
bys id: center UV1 UV2 , pre(cwc_) mean(mwc_)


xtmixed AV cgm_UV1 cgm_UV2 … ||id:
xtmixed AV cwc_UV1 cwc_UV2 … ||id:
```

Assignments

1. Describe an interaction you are interested in
- Graphically, Hypotheses, how to test?

2. Start preparing your own project (feedback on 26 Nov)

3. Read text for next week

Thank you for your Attention!