Multilevel Modelling-course: DAY 3

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

15.11.2021

Schedule for Day 3

- Mediation
- Growth curve models
- Logit models (FYI)
- Your models – your questions – my advise

Schedule for Day 3

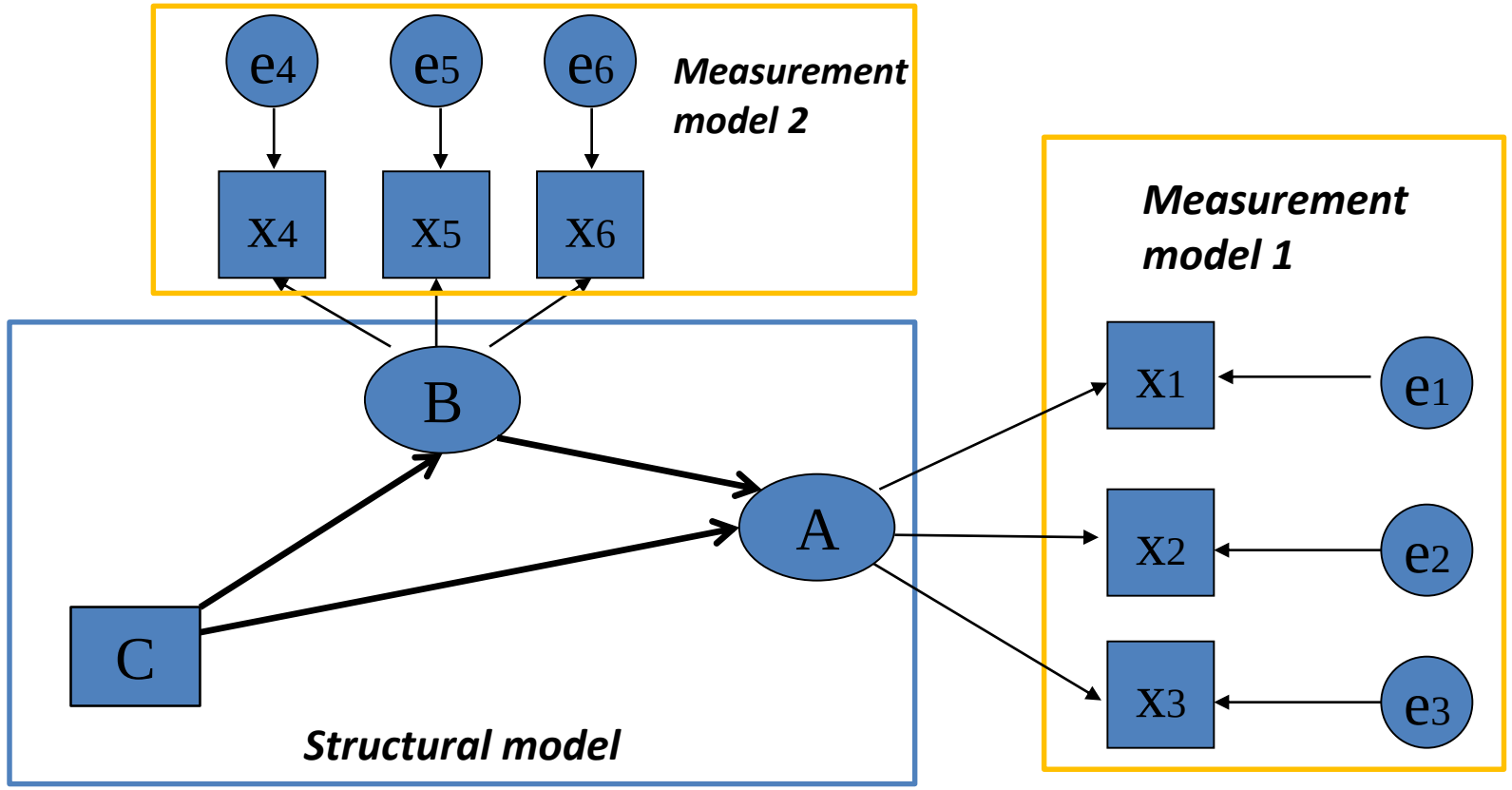
- **Mediation**
- Growth curve models
- Logit models (FYI)
- Your models – your questions – my advise

Structural Equation Models

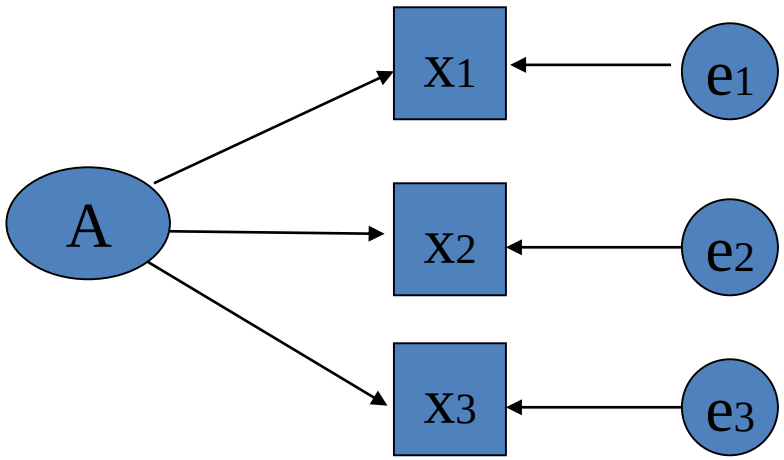
Structural Equation Models (= SEM)

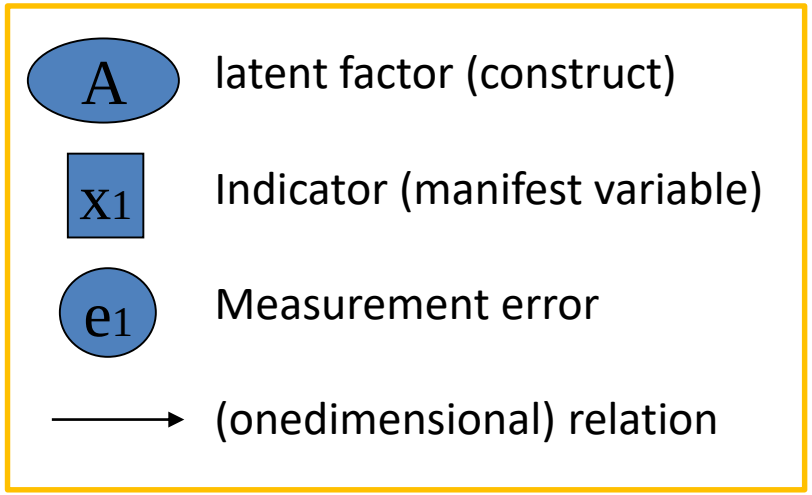
- consist of systems of equations
- can be divided into measurement models and structural models
- direct and indirect effects can be distinguished (mediation)
- distinguish between latent (non-measured) and manifest (measured) variables
- typical graphical representation of the models

Structural Equation Models

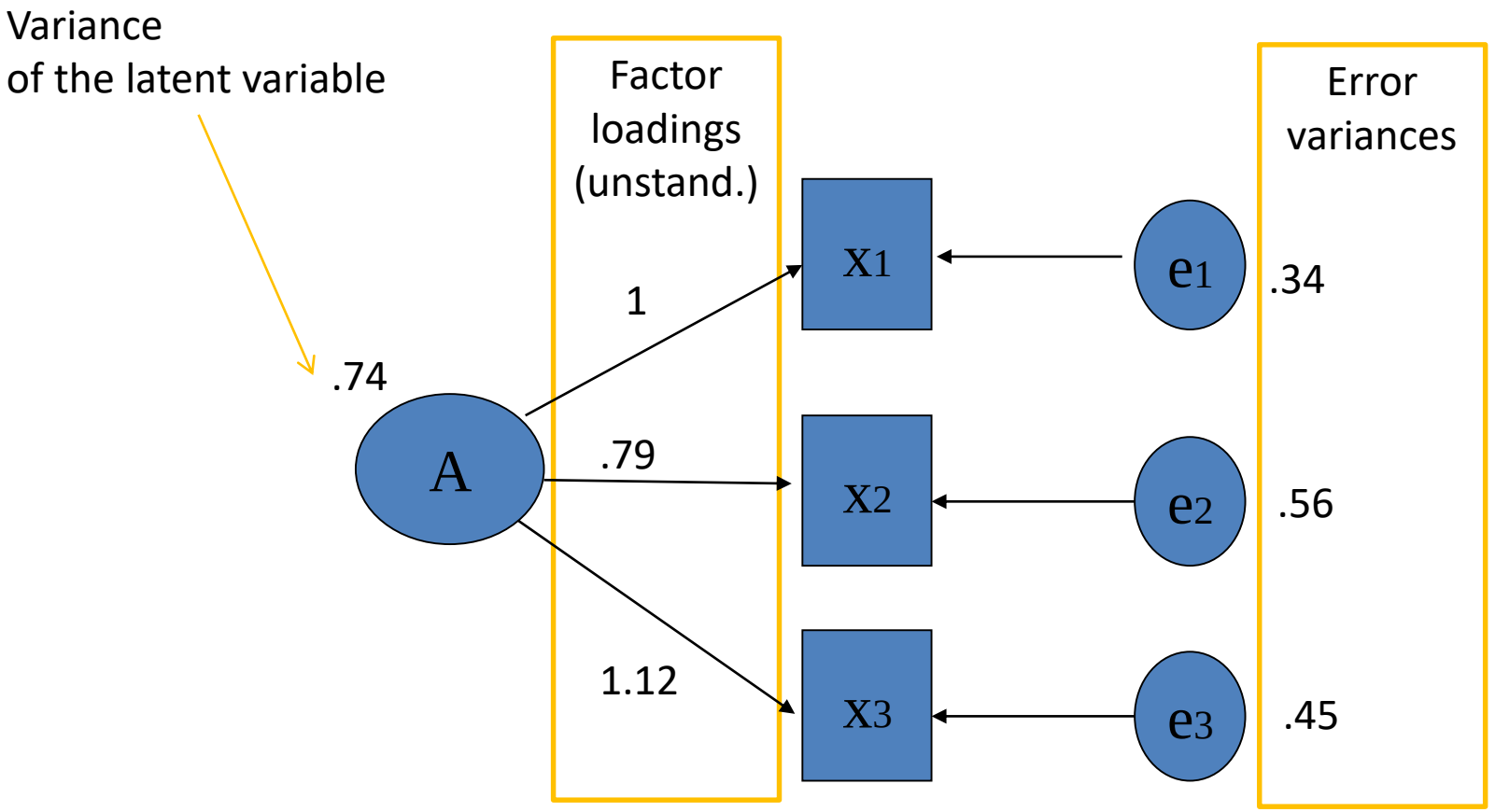


Measurement Model with one Factor

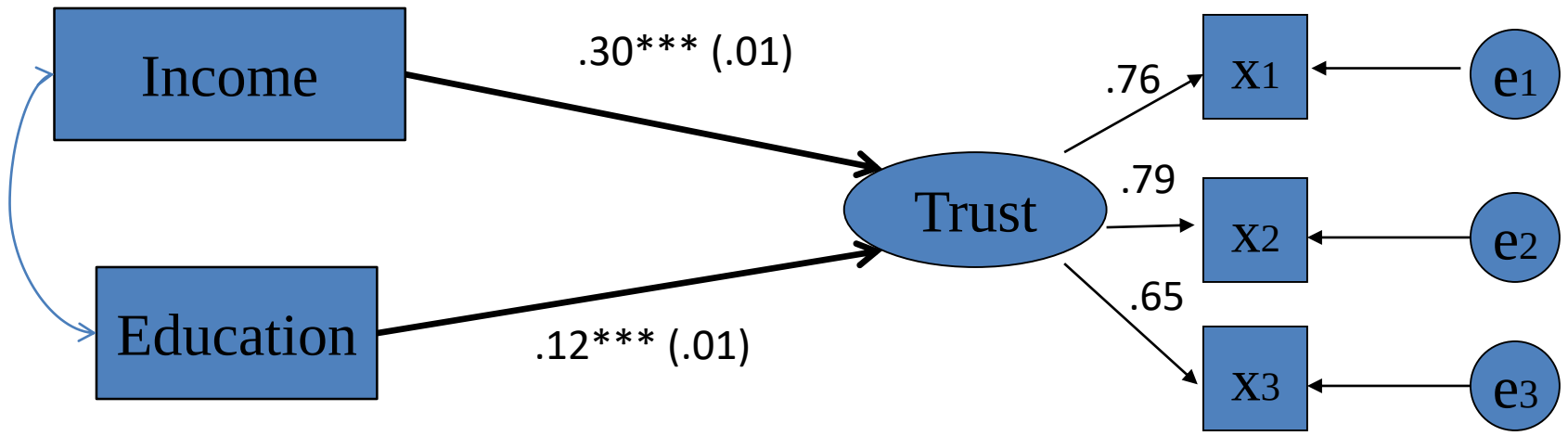




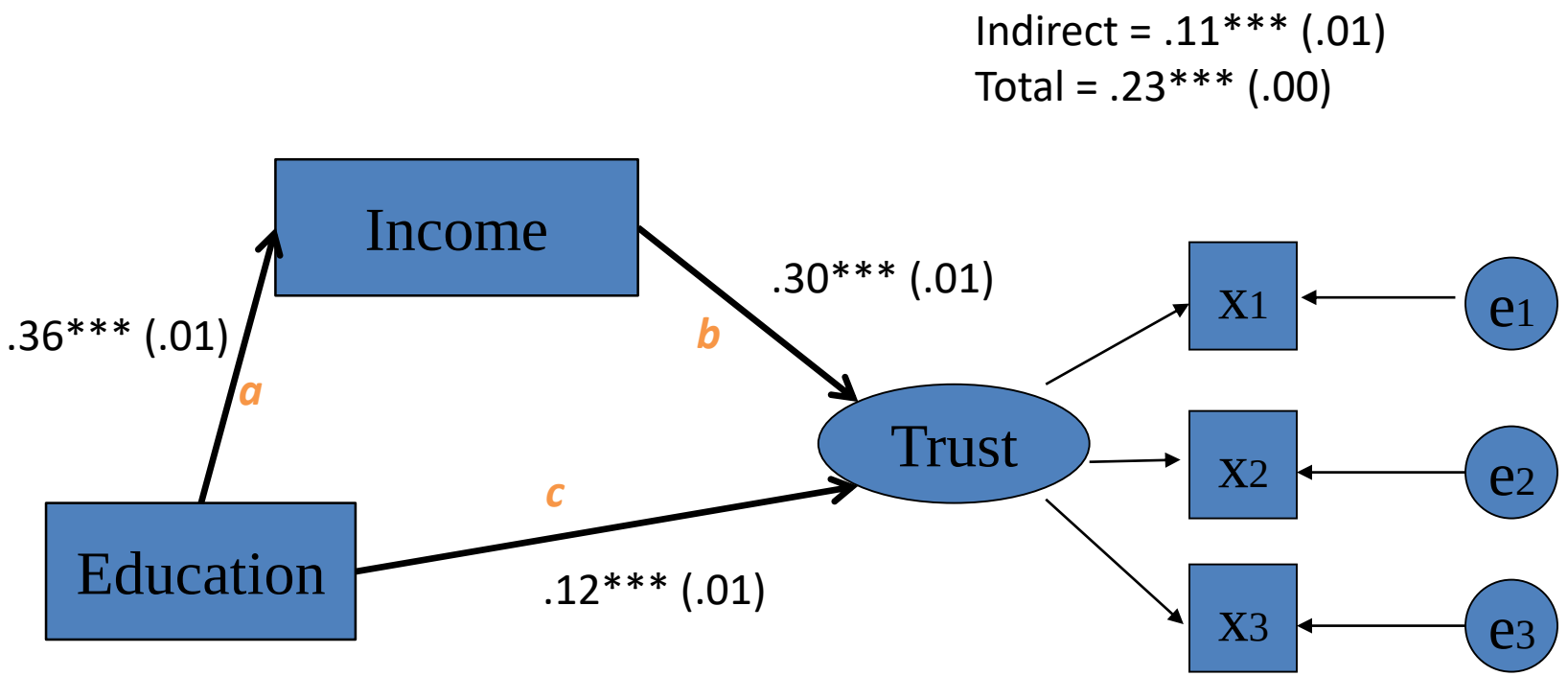
Measurement Model: Parameter (Loadings and variances)



Example 1: Social Trust



Example 2: Social Trust



Indirect effect of education = a * b

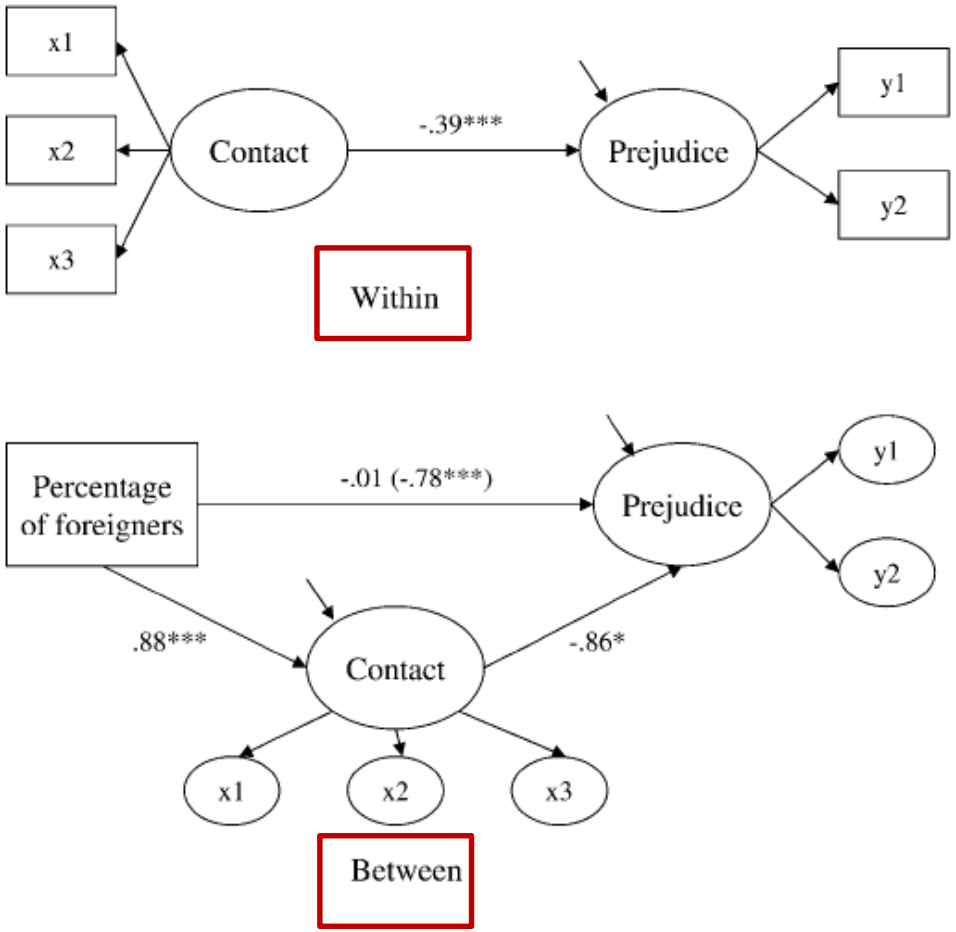
Total effect of education= indirect + direct effect or a * b + c

Stata

```
* Example 1
sem (ppltrst pplfair pplhlp <- Trust) ///
(Trust <- income educyears) , latent(Trust) stand
estat gof , stats(all) // shows fit-indices


* Example 2
sem (ppltrst pplfair pplhlp <- Trust) ///
(Trust <- income educyears) ///
(income <- educyears) , latent(Trust) stand
estat teffects, stand //shows indirect effect
estat gof , stats(all)
```

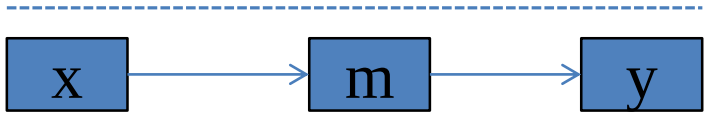
Example Multi-Level SEM



Wagner, U., Christ, O., Pettigrew, T.F., Stellmacher, J., & Wolf, C. (2006). Prejudice and minority proportion: Contact instead of threat effects. *Social Psychology Quarterly*, *69*, 380-390.

Typical Multi-Level SEM-Mediationmodel

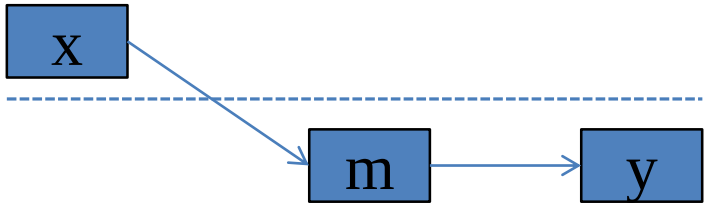
- 1-1-1



BETWEEN

WITHIN

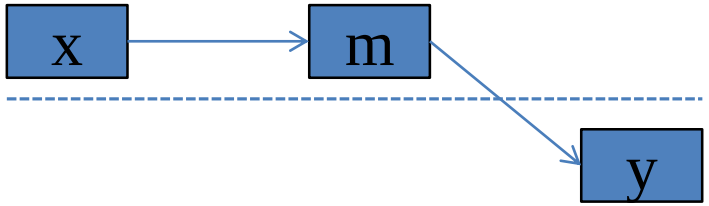
- 2-1-1



BETWEEN

WITHIN

- 2-2-1



BETWEEN

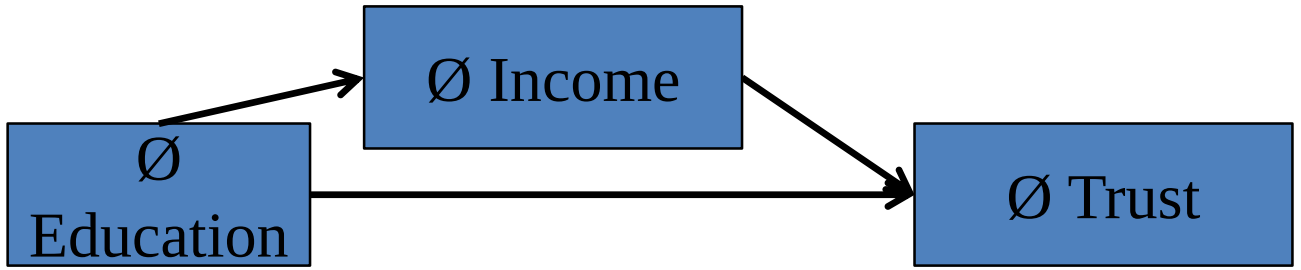
WITHIN

Multi-Level SEM

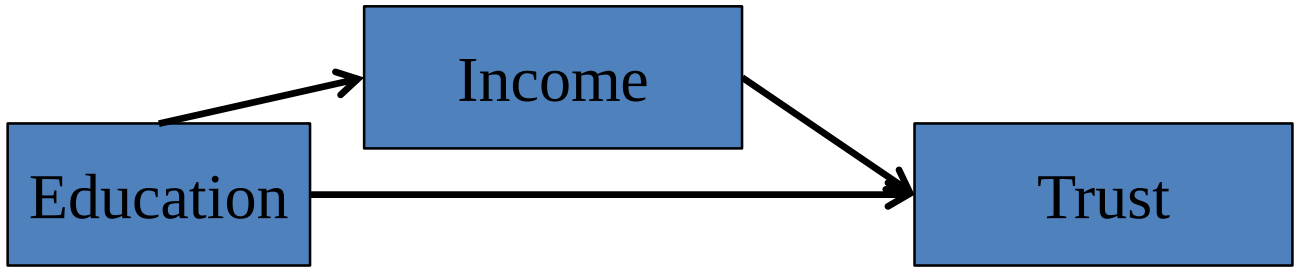
- Without latent variables or mediation, multilevel regression and multilevel SEM produce equivalent results
- Latent measurement models (CFA) only reliable with large number of clusters (> 60)
- With multilevel mediation and level-2 involved, the between-level is one that is interpreted (Preacher et al. 2010)
- Mplus has enormous advantages over Stata in specifying ML-SEM models; Mplus syntax see http://www.quantpsy.org/pubs/syntax_appendix_081311.pdf
- In R: lavaan package mimicks Mplus; Stata: gsem (slow and often does not converge)

Example 3: Social Trust on Individual and Country-Level

BETWEEN



WITHIN

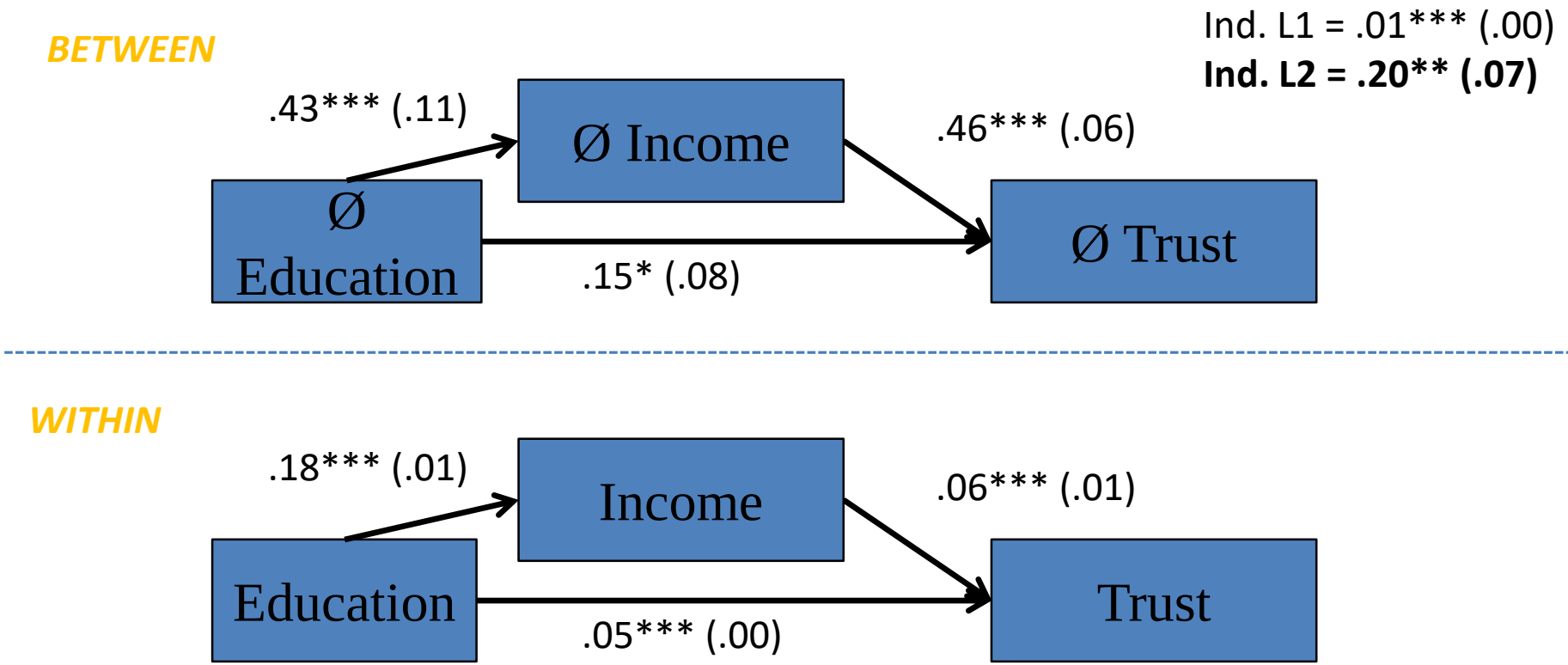


Mplus – Output

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
TRUST ON				
INCOME	0.056	0.011	5.286	0.000
EDUCYEARS	0.046	0.004	11.346	0.000
INCOME ON				
EDUCYEARS	0.177	0.009	20.568	0.000
Residual Variances				
TRUST	3.063	0.157	19.554	0.000
INCOME	3.487	0.193	18.066	0.000
Between Level				
TRUST ON				
INCOME	0.458	0.062	7.443	0.000
EDUCYEARS	0.148	0.076	1.960	0.050
INCOME ON				
EDUCYEARS	0.429	0.113	3.801	0.000
Intercepts				
TRUST	0.669	0.870	0.769	0.442
INCOME	1.022	1.342	0.761	0.447
Residual Variances				
TRUST	0.212	0.048	4.462	0.000
INCOME	1.646	0.614	2.681	0.007
New/Additional Parameters				
ITRUST	0.010	0.002	5.752	0.000
ITRUST2	0.196	0.070	2.814	0.005

Example 3: Social Trust on Individual- and Country – Level



Group session

Come together in groups and find examples for all three types of multilevel mediation:

Predictor-Mediator-Outcome

L1-L1-L1

L2-L1-L1

L2-L2-L1

Schedule for Day 3

- Mediation
- **Growth curve models**
- Logit models (FYI)
- Your models – your questions – my advise

Growth Curve Model

Questions (example DV: Social trust, period 2002-2006)

1. Can a change in social trust be observed in the period 2002-2006 and (b) is there variability in the change over time between respondents?
2. If there is significant variability in change over time, can this variability be explained by specific variables?

- Growth curve models can be estimated as ML regression models as well as ML-SEM models
- Complex, non-linear developments can be taken into account
- Particularly suitable for genuine panel data of individuals
- Attention: Both between- and within-variance are included in the estimates (no separation between the two)

Growth Curve Model

$$y_{ti} = \pi_{0i} + \pi_{1i} t_{ti} + e_{ti}$$

- π_{0i}: trust value of a person at the initial time t_0
- $\pi_{1i} t_{ti}$: Linear increase of the trust value per time point
- e_{ti}: Residual of an observation

Basically a RIRS-model with repeated measurements nested in persons (random intercept) & with time as predictor and random slope

$$\pi_{0i} = \beta_{00} + u_{0i}$$

- β_{00}: Overall average at baseline
- u_{0i}: Deviations of single individuals

Comparable with Random Intercept

$$\pi_{1i} = \beta_{10} + u_{1i}$$

- β_{10}: Average linear increase per time point
- u_{1i}: Deviation speed of increase/decrease for an individual

Comparable with Random Slope

 Overall equation: $y_{ti} = \beta_{00} + \beta_{10} t_{ti} + u_{1i} t_{ti} + u_{0i} + e_{ti}$

Growth Curve Model

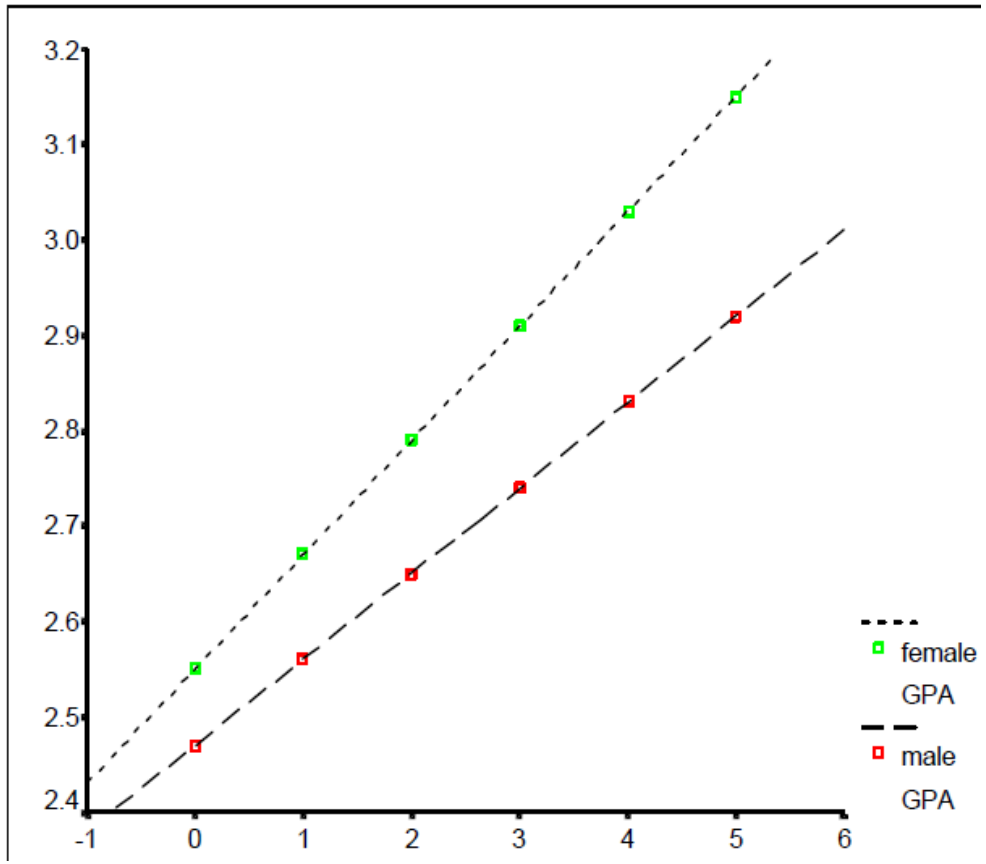
- Additional time-varying and time-constant covariates can be added
- In interaction with time, we want to explain different temporal trajectories with substantial variables (e.g., does gender (Z) explain the found variability in over-time development across individuals (u_{1i}) → cross-level interaction
- Polynomial curves can be added by including quadratic, cubic, … time effects; even dummy variables for time are feasible
- Growth curve models allow for unbalanced panel data

Growth Curve Model

- Example Hox (2010: 90)

$$GPA_{ti} = \beta_{00} + \beta_{10}Time_{ti} + \beta_{20}Job_{ti} + \beta_{01}Sex_i + \beta_{11}Sex_iTime_{ti} + u_{1i}Time_{ti} + u_{0i} + \varepsilon_{ti}$$

Fixed part	**Coeff.**	**S.E.**
Intercept	2.58	(0.09)
Occasion	0.09	(0.01)
Job status	-0.13	(0.02)
GPA highschool	0.09	(0.03)
Gender	0.08	(0.03)
Occasion*Gender	0.03	(0.01)
Random part		
σ²(e)	0.042	(0.002)
σ²(u0)	0.038	(0.010)
σ²(u1)	0.004	(0.001)
σ(u0 u1)	-0.002	(0.001)

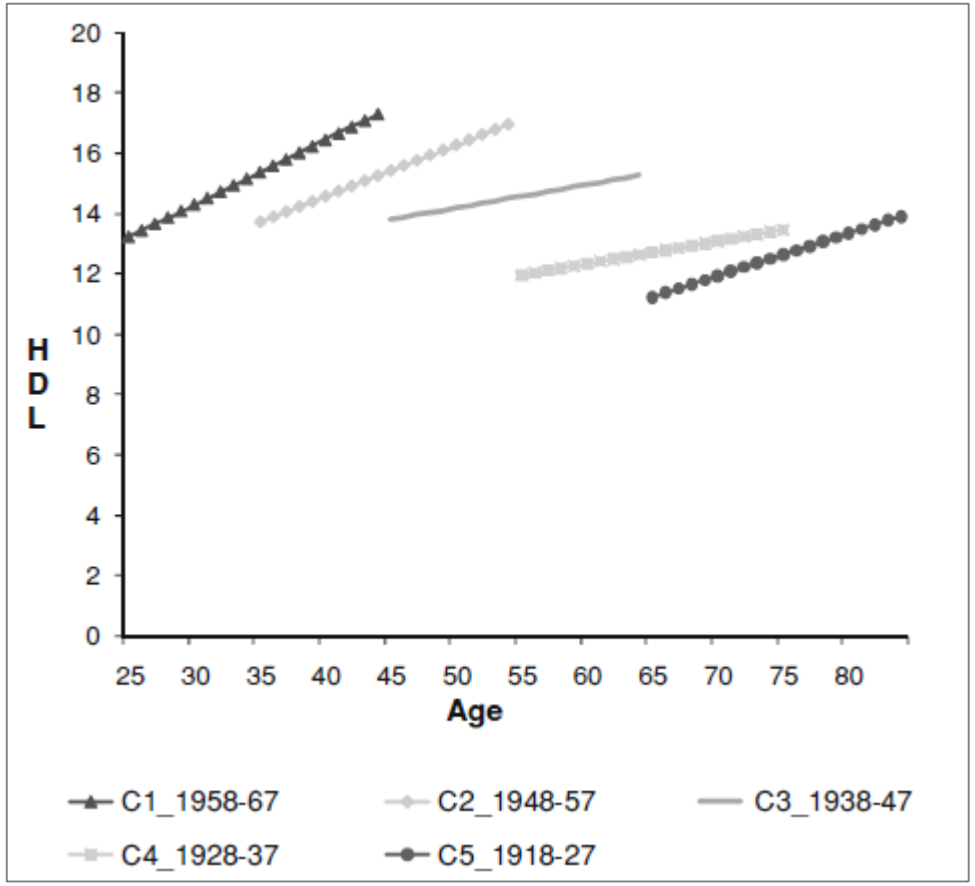


Growth Curve Model

Example: Brault, M.-C., Meuleman, B. & Bracke, P. (2012) Depressive symptoms in the Belgian population: disentangling age and cohort effects. *Journal of Social Psychiatry and Psychiatric Epidemiology*.

Fixed effects		
Initial status		
Intercept y_{00}	13.65	***
C1 1958–1967	4.366	***
C2 1948–1957	1.930	***
C3 1938–1947	REF	
C4 1928–1937	−2.808	***
C5 1918–1927	−5.489	***
Woman	3.360	***
Education	−0.084	*
Married	REF	
Single/widowed	0.248	
Divorced/separated	0.349	
Partner	−1.042	**
Monthly income €	−0.163	
Employed	−0.609	***

Growth rate		
Agec y_{10}	0.054	
Agec2 y_{20}		
Agec × C1_5867	0.165	***
Agec × C2_4857	0.105	**
Agec × C3_3847	REF	
Agec × C4_2837	−0.003	
Agec × C5_1827	0.065	
Variance components (random effects)		
Level 1: within-person	30.315	***
Level 2: in initial status	43.739	***
Level 2: in linear growth	0.087	***
Proportion of variance explained		
Within-person	2.147	6.6%
In initial status	7.708	15.0%
In linear growth	0.008	8.4%



Stata

```
//Gowth Curve Model

mixed y t ||id: t   , cov(un)

// Growth curve model with predictors to explain differences
at baseline(= Random Intercept)

mixed y t x ||id: t   , cov(un)

// Growth curve model with predictors to explain differences
in growth curves (= interaction to explain slope variance)

mixed y t x c.x#c.t ||id: t   , cov(un)
```

Schedule for Day 3

- Mediation
- Growth curve models
- **Logit models (FYI)**
- Your models – your questions – my advise

Logistic Multilevel-Analysis

- Modelling dichotomous dependent variables as a function of individual and contextual explanatory factors $P(y_{ij} = 1 | x_{ij}, z_j)$

- Random Intercept Model
  - Logistic regression function for each country separately

$$\ln\left(\frac{p}{1-p}\right) = \beta_{0j} + \beta_{1j} x_{ij}$$

Average log(odds) in each group

Overall effect of x on the log(odds)

No error term!

Logistic Multilevel-Analysis

- Random Intercept Modell
  - Determination of the group-specific intercepts

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

Average log(odds) across all groups

Difference in logged odds für group j

$$\ln\left(\frac{p}{1-p}\right) = \gamma_{00} + \beta_{1j}x_{ij} + u_{0j}$$

Logistic Multilevel-Analysis

- Residual variance is fixed at Pi^2/3
- ICC is computed as

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \pi^2 / 3}$$

- Fixation leads to rescaling of coefficients in case of model changes; makes comparison of models difficult (calculate marginal effects!)
- Model fit as proportional reduction of variance

$$R_{MZ}^2 = \frac{\sigma_F^2}{\sigma_F^2 + \sigma_{u0}^2 + \pi^2 / 3}$$

σ_F^2 = Residual variance of a linear prediction of the estimated model parameters (see Snijders & Bosker 2012: 306)

Stata

```
//compute ICC
disp  var(_cons) / (var(_cons) + _pi^2 / 3)

//Random Intercept Modell
melogit DV IV1 IV2 … ||id:

//Random Slope Modell
melogit DV IV1 IV2 … ||id: IV1  , cov(un)

margins, dydx(*) predict(mu fixedonly)
```

Schedule for Day 3

- Mediation
- Growth curve models
- Logit models (FYI)
- **Your models – your questions – my advise**

Thank you for your Attention!