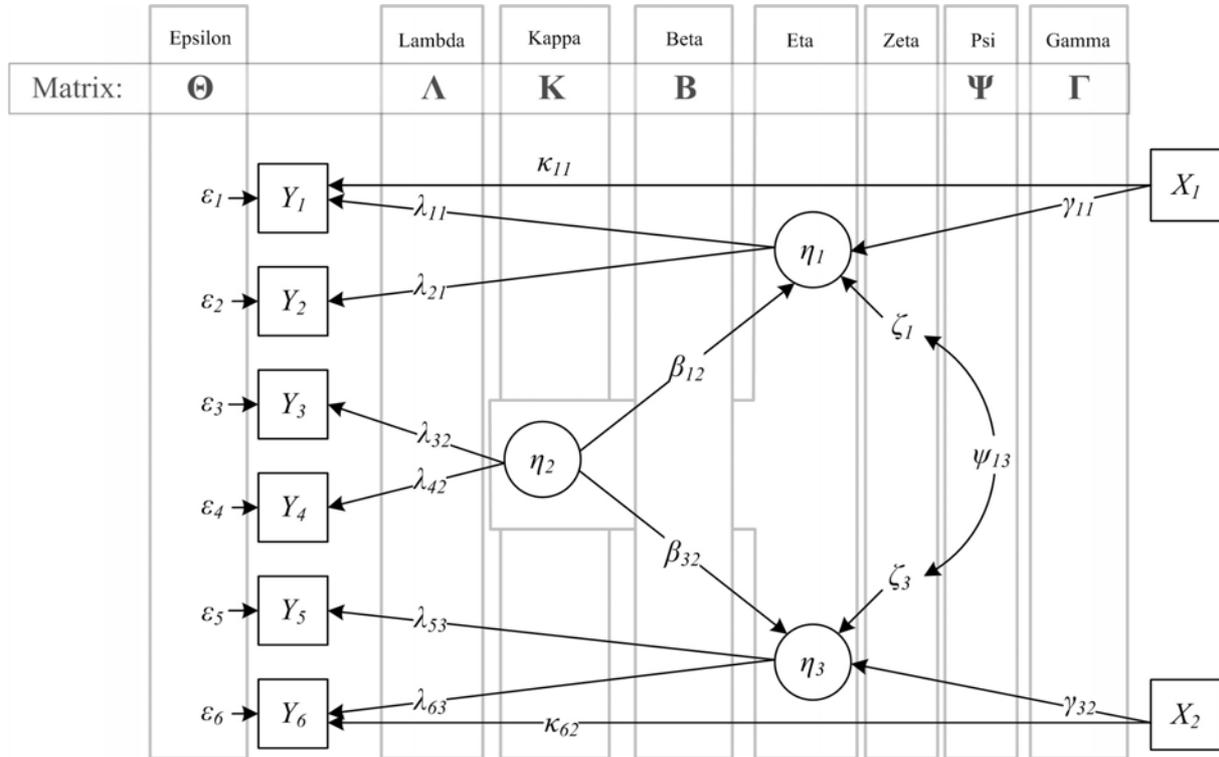


SEM and latent variable modelling with *Mplus*

Pathdiagram:



The general *Mplus*-model for continuous variables Y :

- Measurement model: $Y = v + \Lambda\eta + KX + \varepsilon$
- Structural model: $\eta = \alpha + B\eta + \Gamma X + \zeta$

Implied covariance-structure of the *Mplus*-model for continuous variables Y :

$$\Sigma(\theta) = \left[\Lambda(\mathbf{I} - \mathbf{B})^{-1} \Gamma \Sigma_{xx} \Gamma' (\mathbf{I} - \mathbf{B})^{-1} \Lambda' + \Lambda(\mathbf{I} - \mathbf{B})^{-1} \Psi (\mathbf{I} - \mathbf{B})^{-1} \Lambda' + \Theta \right]$$

Implied mean structure of the *Mplus*-model for continuous variables Y :

$$E(Y) = v + \Lambda(\mathbf{I} - \mathbf{B})^{-1} [\alpha + \Gamma E(X)] + K E(X)$$

$$E(\eta) = (\mathbf{I} - \mathbf{B})^{-1} [\alpha + \Gamma E(X)]$$

Note: In SEM with manifest variables Y which are binary or ordered categorical the model implied mean and covariance structure holds for the latent response variables Y^* , whereby:

$$Y_j = c \Leftrightarrow \tau_c < Y_j^* < \tau_{c+1} \quad \text{for the categories } c = 0, 1, \dots, C; \quad \text{with } \tau_0 = 0 \text{ and } \tau_C = +\infty.$$

Variables:

- X the q exogenous manifest variables
- Y the p endogenous manifest variables
- η the m exogenous and endogenous latent variables

Parameter Arrays:**Arrays for the STRUCTURAL EQUATION PART of the model**

TAU	It contains information regarding thresholds of categorical observed variables
NU	It contains information regarding means or intercepts of continuous observed variables.
LAMBDA	It contains factor loadings, e.g. the regression coefficients regarding the regression of the observed variables Y (or latent response variables Y^*) on the latent variables η .
THETA	It contains the residual variances and covariances of the observed dependent variables or the latent response variables.
ALPHA	The alpha vector contains the means and/or intercepts of the continuous latent variables.
BETA	It contains the regression coefficients for the regressions of continuous latent variables on continuous latent variables.
GAMMA	The gamma matrix contains the regression coefficients for the regressions of continuous latent variables on observed independent variables.
PSI	The psi matrix contains the variances and covariances of the continuous latent variables.
DELTA	Delta is a vector that contains scaling information for the observed dependent variables.

Arrays for the MIXTURE PART of the model

ALPHA (C)	It vector contains the mean or intercept of the categorical latent variables.
LAMBDA (U)	The lambda (u) matrix contains the intercepts of the binary observed variables that are influenced by the categorical latent variables.
TAU (U)	The tau (u) vector contains the thresholds of the categorical observed variables that are influenced by the categorical latent variables.
GAMMA (C)	The gamma (c) matrix contains the regression coefficients for the regressions of the categorical latent variables on observed independent variables.
KAPPA (U)	The kappa (u) matrix contains the regression coefficients for the regressions of the binary observed variables on the observed independent variables.

Arrays for the GROWTH MODELING - MIXTURE PART of the model

ALPHA (F)	The alpha (f) vector contains the means and/or intercepts of the growth factors for the categorical observed variables that are influenced by the categorical latent variables.
LAMBDA (F)	The lambda (f) matrix contains the fixed loadings that describe the growth of the categorical observed variables that are influenced by the categorical latent variables.

Commands and options

TITLE:	title for the analysis		
DATA:	FILE IS	file name;	
	TYPE IS	INDIVIDUAL; (default)	Datamatrix where rows represent observations and columns represent variables
		COVARIANCE;	A lower triangular covariance matrix
		CORRELATION;	A lower triangular correlation matrix
		FULLCOV;	A full triangular covariance matrix
		FULLCORR;	A full triangular correlation matrix
		MEANS;	Vector of means
	STDEVIATIONS;	Vector of standard deviations	
	NOBSERVATIONS ARE	number of observations;	
	NGROUPS =	number of groups;	
VARIABLE:	NAMES ARE	names of variables in the data set;	
	USEOBSERVATIONS ARE	conditional statement to select observations;	default: all observations in data set
	USEVARIABLES ARE	names of analysis variables;	default: all variables in NAMES
	MISSING ARE	variable (#);	
		;	
		*	
		BLANK;	
	CENSORED ARE	names, censoring type, and inflation status for censored dependent variables;	
	CATEGORICAL ARE	names of binary and ordered categorical (ordinal) dependent variables;	
	NOMINAL ARE	names of unordered categorical (nominal) dependent variables;	
	COUNT ARE	names and inflation status for count variables;	
	GROUPING IS	name of grouping variable (labels);	
ANALYSIS:	TYPE =	GENERAL; (default)	SEM with latent and manifest variables
		BASIC;	Computes sample statistics and other descriptive information
		MEANSTRUCTURE;	Analysis of means, thresholds and intercepts
		MISSING;	Analysis of data containing missing values
		HI;	Estimation of unrestricted mean and covariance model with TYPE = MISSING
		RANDOM;	Models with both random intercepts and random slopes
		COMPLEX;	Models with non-independence of observations due to cluster sampling
		MIXTURE;	Models with a combination of continuous and categorical latent variables
		BASIC;	Computes sample statistics and other descriptive information
		MISSING;	Analysis of data containing missing values
		RANDOM;	Models with both random intercepts and random slopes
		COMPLEX;	Models with non-independence of observations due to cluster sampling
		TWOLEVEL;	Multilevel modeling
		BASIC;	Computes sample statistics and other descriptive information
		MISSING;	Analysis of data containing missing values
		HI;	Estimation of unrestricted mean and covariance model with TYPE = MISSING
		RANDOM;	Models with both random intercepts and random slopes
		MIXTURE;	Models with a combination of continuous and categorical latent variables
		EFA	Exploratory factor analysis
		BASIC;	Computes sample statistics and other descriptive information
	MISSING;	Analysis of data containing missing values	
	ESTIMATOR = (depends on analysis type)	ML;	Maximum Likelihood
		MLM;	Maximum Likelihood – robust to non-normality (Satorra-Bentler χ^2)
		MLMV;	ML with mean and variance adjusted χ^2 -statistic – robust to non-normality
		MLR;	ML – robust to non-normality and non-independence of observations with TYPE=COMPLEX (Yuan-Bentler T^2 - Test statistic)
		MLF;	ML with standard errors approximated by first order derivatives
		MUML;	Muthén's limited information parameter estimates
		WLS;	Weighted Least Square estimator
		WLSM;	Weighted Least Square estimator using a diagonal weight matrix and with mean - adjusted χ^2 -statistic
		WLSMV;	Weighted Least Square estimator using a diagonal weight matrix and with mean - and variance-adjusted χ^2 -statistic
		GLS;	Generalized Least Square estimator
	ULS;	Unweighted Least square estimator use a full weight matrix	
	PARAMETERIZATION = (depends on analysis type)	DELTA;	Parameterization in SEM with ordered categorical manifest variables
		THETA;	
		LOGIT;	Parameterization in mixture modeling with continuous and ordered categorical manifest and latent variables
	INFORMATION = (depends on analysis type)	LOGLINEAR;	
		OBSERVED;	Estimates information matrix using observed second-order derivatives
		EXPECTED;	Estimates information matrix using expected second-order derivatives
	COMBINATION;	Estimates information matrix using a combination of expected and observed second-order derivatives	
	BOOTSTRAP =	number of bootstrap draws;	Number of bootstrap for computation of standard errors and confidence intervals
	DIFFTEST =	file name;	Option is used for specification of the H_0 -model. The file contains test statistic of the H_1 -model to obtain the correct χ^2 -difference of the likelihood ratio test when MLMV or WLSMV is used
	MODEL:	BY	short for measured by -- defines latent variables (example: f1 BY y1-y5;)
		ON	short for regressed on -- defines regression relationships (example: f1 ON x1-x9;)
		WITH	short for correlated with -- defines correlational relationships (example: f1 WITH f2;)
		PWITH	short for correlated with -- defines paired correlational relationships (example: f1 f2 f3 PWITH f4 f5 f6;)
		list of variables;	refers to variances and residual variances (example: f1 y1-y9;)
		[list of variables];	refers to means, intercepts, thresholds (example: [f1, y1-y9];)
		*	frees a parameter at a default value or a specific starting value (example: y1* y2*.5;)
		@	fixes a parameter at a default value or a specific value (example: y1@ y2@0;)
		(number)	constrains parameters to be equal (example: f1 ON x1 (1); f2 ON x2 (1);)
variable\$number		label for the threshold of a variable	
variable#number		label for nominal observed or categorical latent variable	
variable#1		label for censored or count inflation variable	
variable#number		label for baseline hazard parameters	
variable#number		label for a latent class	
(name)		label for a parameter (used for equality constraints, linear- and non-linear constraints)	
{list of variables};		refers to scale factors (example: {y1-y9};)	
		names and defines random effect variables (example: s y1 ON x1;)	
AT		short for measured at -- defines random effect variables (example: s y1-y4 AT t1-t4;)	
XWITH		defines interactions between variables	
MODEL INDIRECT:		describes the relationships for which indirect and total effects are requested	

	IND	describes a specific indirect effect or a set of indirect effects
	VIA	describes a set of indirect effects that includes specific mediators
	MODEL CONSTRAINT:	describes linear and non-linear constraints on parameters
	NEW	assigns labels to parameters not in the analysis model
	MODEL TEST:	describes restrictions on the analysis model for the Wald test
	MODEL:	describes the analysis model
	MODEL label:	describes the group-specific model in multiple group analysis and the model for each categorical latent variable and combinations of categorical latent variables in mixture modelling
	MODEL: %OVERALL% %class label%	describes the overall part of a mixture model describes the class-specific part of a mixture model
	MODEL: %WITHIN% %BETWEEN%	describes the within part of a two-level model describes the between part of a two-level model
	MODEL:	describes the analysis model
	MODEL label:	describes the group-specific model in multiple group analysis and the model for each categorical latent variable and combinations of categorical latent variables in mixture modeling
	MODEL: %OVERALL% %class label%	describes the overall part of a mixture model describes the class-specific part of a mixture model
	MODEL: %WITHIN% %BETWEEN%	describes the within part of a two-level model describes the between part of a two-level model
	MODEL:	describes the analysis model
	MODEL label:	describes the group-specific model in multiple group analysis and the model for each categorical latent variable and combinations of categorical latent variables in mixture modeling
	MODEL: %OVERALL% %class label%	describes the overall part of a mixture model describes the class-specific part of a mixture model
	MODEL: %WITHIN% %BETWEEN%	describes the within part of a two-level model describes the between part of a two-level model
	MODEL:	describes the analysis model
	MODEL label:	describes the group-specific model in multiple group analysis and the model for each categorical latent variable and combinations of categorical latent variables in mixture modelling
	MODEL: %OVERALL% %class label%	describes the overall part of a mixture model describes the class-specific part of a mixture model
	MODEL: %WITHIN% %BETWEEN%	describes the within part of a two-level model describes the between part of a two-level model
OUTPUT:	SAMPSTAT;	Option is used to request sample statistics for the data being analyzed (observed covariances, correlations, means, thresholds)
	STANDARDIZED;	Option is used to request standardized coefficients
	RESIDUAL;	Option is used to request model estimated means/intercepts/thresholds and/or model estimated covariances/correlations and the concerning residuals of these parameters.
	MODINDICES (minimum chi-square);	Option is used to request modification indices, expected parameter change and standardized expected parameter change
	CINTERVAL; CINTERVAL (SYMMETRIC); CINTERVAL (BOOTSTRAP); CINTERVAL (BCBOOTSTRAP);	Option is used to request confidence intervals for parameter estimates, indirect effects and standardized indirect effects. The option has three settings: SYMMETRIC, BOOTSTRAP and BCBOOTSTRAP! Symmetric computes the confidence intervals via standard errors! Confidence intervals determined on the bootstrap method take non-normality of the parameter estimate distribution into account! BCBOOTSTRAP refers to bias corrected bootstrap method.
	NOCHISQUARE;	Option is used to request that the χ^2 -fit statistic not be computed.
	NOSERROR;	Option is used to request that the standard errors statistic not be computed.
	PATTERNS;	Option is used to request a summary of the missing data patterns.
	FSCOEFFICIENT;	Option is used to request factor score coefficients and a factor score posterior covariance matrix.
	FSDETERMINACY;	Option is used to request factor score determinacy value for each factor in the model.
	TECH1;	Option is used to request the arrays containing parameter specification and starting values of all free parameters in the model.
	TECH2;	Option is used to request parameter derivatives.
	TECH3;	Option is used to request estimated covariance and correlation matrices of the parameter estimates.
	TECH4;	Option is used to request estimated means, covariances and correlations for the latent variables in the model.
	TECH5;	Option is used to request the optimization history in estimating the model.
	TECH6;	Option is used to request the optimization in estimating the sample statistics for categorical observed variables.
	TECH7;	Option is used in conjunction with TYPE=MIXTURE to request sample statistics for each class.
	TECH8;	Option is used in conjunction with TYPE=MIXTURE to request the optimization history in estimating the model.
	TECH9;	Option is used in conjunction with MONTECARLO to request error messages related to convergence of each replication.
	TECH10;	Option is used in conjunction with TYPE=MIXTURE to request uni- and bivariate model fit information of the categorical dependent variables.
	TECH11;	Option is used in conjunction with TYPE=MIXTURE to request the Lo-Mendell-Rubin likelihood ration test of model fit.
	TECH12;	Option is used in conjunction with TYPE=MIXTURE to request residuals of observed versus model estimated means, variances, covariances, univariate skewness and kurtosis.
	TECH13;	Option is used in conjunction with TYPE=MIXTURE to request
SAVEDATA:	FILE IS file name;	Option is used to specify the name of the ASCII file in which the individual-level data used in the analysis will be saved.
	SAMPLE IS file name;	Option is used to specify the name of the ASCII file in which the sample statistics will be saved.
	SIGB IS file name;	Option is used to specify the name of the ASCII file in which the sigma between covariance/correlation matrix (if TYPE=TWOLEVEL) will be saved.
	RESULTS ARE file name;	Option is used to specify the name of the ASCII file in which the results of the analysis will be saved.
	ESTIMATES ARE file name;	Option is used to specify the name of the ASCII file in which the parameter estimates will be saved.
	DIFFTEST IS file name;	Option is required in the input file of the H ₁ -model when the likelihood ratio test is applied and the MLMV- or WLSMV-estimator is used. The file contains test statistic of the H ₁ -model, which is needed in the subsequent computation of the correct χ^2 -difference between the H ₁ - and H ₀ -model!
	TECH3 IS file name;	Option is used to specify the name of the ASCII file in which the covariance and correlation matrices of the parameter estimates will be saved.
	TECH4 IS file name;	Option is used to specify the name of the ASCII file in which the estimated means, covariances and correlations for the latent variables will be saved.
	FORMAT IS format statement; FREE; (default)	Option is used to specify the format the in which the analysis data will be saved.
	TYPE IS	COVARIANCE; CORRELATION; Option is used to specify the type of matrix to be saved. It can be used in conjunction with the SAMPLE and SIGB options to override the default matrix that is saved.
	RECORDLENGTH IS characters per record;	Option is used to specify number of characters per record in the file to which the analysis data are saved. The default an maximum record length is 1000.
	SAVE =	FSCORES; CPROBABILITIES; Factor scores are saved along with the other analysis variables. Option is used in conjunction with TYPE=MIXTURE and SAVE=CPROBABILITIES. Individual posterior class probabilities are saved along with the other analysis variables.