Lineaire structurele modellen

College 4: Inleiding LISREL

Harry Ganzeboom
MA-cursus
23 februari 2009

LISREL

- SEM: Simultaneous [Structural] Equations Model:
  - A system of linear equations ('causal model')
  - Coefficients are calculated all at once.
- LISREL is the simultaneous estimation of:
  - A structural model ('causal model')
  - A measurement model ('factor analysis').

General procedure

- Conceive of the world as a correlation (covariance) matrix.
- (that represents a multivariate normal distribution – this is an untested assumption in LISREL)
- Think of a set of linear equations in a causal / measurement model that could (re)produce this correlation matrix.
- Write up these equations from a causal diagram with numbered variables.
- Translate the causal diagram into lisrel's notation.
- F5

Advantages of SEM's

- Full information procedures: you use all the available evidence to estimate coefficients. This increases power in unsaturated models (not all possible effects are estimated).
- Correction for attenuation by random and systematic measurement error.
- It makes you think about measurement error.
- Significance test for all coefficients (also factor loadings)
- Adequate treatment of missing values possible.
- Constrained estimation (restrictions on coefficients) possible.

Disadvantages of LISREL

- Stand-alone program, cumbersome to handle.
- Complicated to get it right, and not only when you are a novice.
- The world may be a correlation matrix, but it is certainly not true that these correlations derive from a multivariate normal distribution. Assumption is not plausible.
- Interaction effects are hard to handle.

Elements of a LISREL syntax

- TI: Title
- DA: Data
- LA: Labels for the input data
- SE: Selection and order of variables
- MO: Model specifications
- LE / LK: Labels for the latent variables
- ST: Specifying and starting values for the fixed element
- FR: Specifying free elements
- FI: Fixing elements
- EQ: Equalizing elements
- OU: Output specifications

Variables in LISREL

- Latent variables in two kinds:
  - Exogenous: KSI
  - Endogenous: ETA
- Observed variables:
  - Exogenous: X
  - Endogenous: Y
- However, it is possible to avoid the specification of exogenous variables (KSI, X) altogether.

Elements of the structural model

- The general format of the structural model is a system of regression equations:
  - **ETA1 = GA1*KSI1+GA2*KSI2 + … +BE1*ETA2 + BE2*ETA3 + … + zeta**
- Note:
  - There is no intercept
  - There are two sets of independent variables: X (exogenous) en (other) Y's (endogenous), each with a set of regression coefficients: GAMMA and BETA..
  - The error term is called ZETA.

Elements of the measurement model

- Y = LY*ETA + epsilon
- X = LX*KSI + delta
- ETA and Y are connected by LY
- KSI and X are connected by LX
- The unexplained (co)variances go into TE (and TD).
- If there is no multiple measurement: LY=1, LX=1, TD=0, TE=0.

Variances-covariances

- The equations imply an square, symmetric matrix of explained and unexplained (co)variances for all latent and observed residuals (zeta, theta, epsilon).
- The matrix of (un)explained (co)variances for the latent variables is called PSI for ETA (and PHI for KSI). TE and TD refer to the residual (co)variances of the observed variables.
- The diagonal elements of the PS, PH, TE and TD contain the (fitted) unexplained variance, the off-diagonal element the (fitted) [residual] covariances.

Fitted covariance matrix

- All variance/covariance matrices together imply an expected variance/covariance matrix for the observed variables.
- Estimation makes the expected matrix as close as possible to the observed matrix.
- Fit statistics measure and test the distance between observed and expected (co)variances.
- The distance is expressed in the residuals: these can guide model adaptation.

TITLE

- The title is optional

DATA

- DA has three important elements
  - NI: Number of input variables
  - NOBS: Number of observations
  - MA=KM specifies that the correlation matrix needs to be analyzed.
- Note that there is single NOBS, so no provisions for missing values.
- If you submit a pairwise computed correlation matrix, take for NOBS the median N.
- NG=1. Specifies that one group needs to be read / analyzed.

KM

- After KM FU or KM SY follows the correlation matrix as obtained in SPSS.
- This matrix can be copied from SPSS output (use factor analysis) using Excel, Word and Notepad.
- The matrix can also be read from a file.
- If you want to analyze covariances, read in an SD vector as well.

LA

- LA gives labels to the observed variables.
- End with a slash: /.

SELECT

- SE makes it possible to select and reorder the variables.
- If you want to analyze less variables than are in the correlation matrix, end with /.
- Elements on SE can be either variable numbers or variable labels (as specified on LA).

MODEL

- MO contains specifications for:
  - NX, NK: number of X and KSI variables
  - NY, NE: number of Y and ETA variables
  - GA, BE: FU, FI: coefficient matrices for structural relationships
  - LY, LX: FU, FI: coefficient matrices for measurement relationships
  - PS, PH: SY, FI: (co)variances for latent residuals.
  - TE, TD: SY, FI: (co)variances for observed residuals.
- See Kelloway, Table 4.1 for more options.

LE, LK

- LE and LK specify labels for ETA, respectively KSI.
- Note that we do need as many labels as variables!

ST / FR

- ST (or VA) gives which elements are to be fixed at a certain value (1.0).
- In a measurement model at least one of the measurement relationships needs to be fixed.
- Example: ST 1.0 LY 1 1.
- FR specifies which elements need to be estimated.
- Examples: FR LY 2 1 BE 2 1 PS 3 3 TE 5 5.

FI

- Free elements can be fixed again by the FI statement.
- Example: FI LY 3 1.
- You only need to use this on formerly freed elements.

EQ

- With EQ with can obtain constrained estimation.
- E.g.: EQ BE 3 1 BE 3 2
- This will often boost statistical power.
- However you need theory behind it.

PD

- The PD (=PathDiagram) produces a path diagram with coefficients and elementary fit information.
- The pathdiagram is not produced when there are errors in the syntax or when the job has not run due to other problems.
- Lisrel also has a graphical interface, using this path diagram. I do not know how to use it.

OUTPUT

- Important options for OU:
  - ND=3 number of decimals
  - SE TV: standard errors / T-values
  - SS standardized solution
  - SC completely standardized solution
- See Kelloway Chapter 4.

! Comments

- You can insert a comment line using an exclamation mark. (cfr. * in SPSS).
- This is also the way to make different versions of a syntax file.