Slide 1

Factor Analysis and Its Extensions

Karl G Jöreskog

Uppsala University

Alternative Titles

Factor Analysis at 100: The Last 50 Years

Factor Analysis: 50 Years in 50 minutes

Slide 2

Uppsala Symposium 1953

My history of factor analysis begins just about 50 years ago, or 1953, at the *Uppsala Symposium on Psychological Factor Analysis*. I was in high school then fully unaware of what was going on in Uppsala and I had no idea what factor analysis was. This symposium was hosted by Herman Wold, professor and chair of statistics at Uppsala University. Wold had met with Louis and Thelma Thurstone in Stockholm and he was inspired by their work on factor analysis. There were some prominent people in this symposium including Maurice Bartlett, D.N.Lawley, Georg Rash, and Peter Whittle. The Uppsala Symposium is of minor importance in the history of factor analysis but it had a great consequence for me for six years later Wold suggested that I do a dissertation on factor analysis.

Karl G Jöreskog Factor Analysis and Its Extensions

Slide 3

Communalities

Many papers on factor analysis at that time focused on the question: What numbers should be put in the diagonal of $\mathbf{R}$ to make this approximately equal to $\mathbf{\Lambda\Lambda}'$, where $\mathbf{\Lambda}$ is a $p \times k$ matrix of factor loadings?

$$\mathbf{R}_c \approx \mathbf{\Lambda\Lambda}' \tag{1}$$

The numbers in the diagonal of $\mathbf{R_c}$ are called communalities. Guttman (1956) showed that the squared multiple correlation R_i^2 in the regression of the $i$the variable on all the other variables is a lower bound for the communality of the ith variable:

$$c_i^2 \geq R_i^2 \tag{2}$$

Slide 4

Uniqueness

The counterpart of the communality c_i^2 is the uniqueness $u_i^2 = 1 - c_i^2$. Hence, (2) is equivalent to

$$u_i^2 \leq 1 - R_i^2 = 1/r^{ii} \tag{3}$$

In my dissertation I therefore suggested that

$$u_i^2 = \theta/r^{ii}\ , \tag{4}$$

where θ is a parameter to be estimated.

Slide 5

Statistical Formultion

Much of the discussion in the 50's were procedures for choosing communalities and estimating factor loadings. There was a need for a statistical formulation. So in my dissertation, I suggested that one could estimate the covariance matrix $\mathbf{\Sigma}$ subject to the constraint

$$\mathbf{\Sigma} = \mathbf{\Lambda\Lambda}' + \theta(diag\mathbf{\Sigma}^{-1})^{-1} \tag{5}$$

and I investigated a simple non-iterative procedure for estimating $\mathbf{\Lambda}$ and θ from the sample covariance matrix $\mathbf{S}$. Later I developed a maximum likelihood method for estimating model (5).

Slide 6

Maximum Likelihood Factor Analysis

Factor analysis as a statistical method was formulated already by Lawley (1940), see also Lawley & Maxwell (1963, 1971) and it was further developed by Anderson & Rubin (1956). However, as late as the mid 60's there was no good way for computing the estimates. Using well established notation, the problem is to minimize the function

$$F_{ML}(\mathbf{\Lambda}, \mathbf{\Psi}) = \log \|\mathbf{\Sigma}\| + tr(\mathbf{S\Sigma}^{-1}) - \log \|\mathbf{S}\| - p\ , \tag{6}$$

where

$$\mathbf{\Sigma} = \mathbf{\Lambda\Lambda}' + \mathbf{\Psi}^2\ , \tag{7}$$

and $\mathbf{\Psi}^2$ is the diagonal matrix of error variances.

Slide 7

Heywood Cases

Jöreskog (1967) solved this problem by focusing on the concentrated fit function

$$f(\mathbf{\Psi}) = \min_{\mathbf{\Lambda}} F(\mathbf{\Lambda}, \mathbf{\Psi})\ , \tag{8}$$

which could be minimized numerically. If one or more of the ψ_i^2 gets close to zero, this procedure becomes unstable. Jöreskog (1977) therefore developed this procedure further by reparameterizing

$$\theta_i = \ln \psi_i^2\ , \quad \psi_i = +\sqrt{(e^{\theta_i})} \tag{9}$$

This leads to a very fast and efficient algorithm, the use of which has been very successful.

Slide 8

Other Fit Functions

Unweighted Least Squares(ULS)

$$F_{ULS}(\mathbf{\Lambda}, \mathbf{\Psi}) = \frac{1}{2}tr[(\mathbf{S} - \mathbf{\Sigma})^2] \tag{10}$$

Generalized Least Squares(GLS)

$$F_{GLS}(\mathbf{\Lambda}, \mathbf{\Psi}) = \frac{1}{2}tr[(\mathbf{I} - \mathbf{S}^{-1}\mathbf{\Sigma})^2] \tag{11}$$

Each of these fit functions can also be minimized by minimizing the corresponding concentrated fit function (8).

Slide 9

Weighted Least Squares

$$F_V(\mathbf{\Lambda}, \mathbf{\Psi}) = \frac{1}{2} tr[(\mathbf{S} - \mathbf{\Sigma})\mathbf{V}]^2 \quad (12)$$

$$\text{ULS}: \ \mathbf{V} = \mathbf{I} \quad (13)$$

$$\text{GLS}: \ \mathbf{V} = \mathbf{S}^{-1} \quad (14)$$

$$\text{ML}: \ \mathbf{V} = \hat{\mathbf{\Sigma}}^{-1} \quad (15)$$

ML = Iteratively Reweighted Least Squares

Slide 10

Exploratory Factor Analysis

Exploratory factor analysis is a technique often used to detect and assess latent sources of variation and covariation in observed measurements. It is widely recognized that exploratory factor analysis can be quite useful in the early stages of experimentation or test development. Thurstone's (1938) primary mental abilities, French's (1951) factors in aptitude and achievement tests and Guilford's (1956) structure of intelligence are good examples of this. The results of an exploratory factor analysis may have heuristic and suggestive value and may generate hypotheses which are capable of more objective testing by other multivariate methods.

Karl G Jöreskog Factor Analysis and Its Extensions

Slide 11

As more knowledge is gained about the nature of social and psychological measurements, however, exploratory factor analysis may not be a useful tool and may even become a hindrance.

Most studies are to some extent both exploratory and confirmatory since they involve some variables of known and other variables of unknown composition. The former should be chosen with great care in order that as much information as possible about the latter may be extracted. It is highly desirable that a hypothesis which has been suggested by mainly exploratory procedures should subsequently be confirmed, or disproved, by obtaining new data and subjecting these to more rigorous statistical techniques.

Slide 12

The basic idea of factor analysis is the following. For a given set of response variables $x_1, \ldots, x_p$ one wants to find a set of underlying latent factors $\xi_1, \ldots, \xi_k$, fewer in number than the observed variables. These latent factors are supposed to account for the intercorrelations of the response variables in the sense that when the factors are partialed out from the observed variables, there should no longer remain any correlations between these. If both the observed response variables and the latent factors are measured in deviations from the mean, this leads to the model:

$$x_i = \lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \cdots + \lambda_{in}\xi_k + \delta_i \ , \quad (16)$$

where δ_i, the unique part of x_i, is assumed to be uncorrelated with $\xi_1, \xi_2, \ldots, \xi_k$ and with δ_j for $j \neq i$. In matrix notation (16) is

$$\mathbf{x} = \mathbf{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta} \quad (17)$$

Slide 13

The unique part δ_i consists of two components: a specific factor s_i and a pure random measurement error e_i. These are indistinguishable, unless the measurements x_i are designed in such a way that they can be separately identified (panel designs and multitrait-multimethod designs). The term δ_i is often called the *measurement error* in x_i even though it is widely recognized that this term may also contain a specific factor as stated above.

Slide 14

Rotation

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2 , \tag{18}$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}^2$ are the covariance matrices of $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$, respectively.

Let $\mathbf{T}$ be an arbitrary non-singular matrix of order $k \times k$ and let

$$\boldsymbol{\xi}^* = \mathbf{T}\boldsymbol{\xi} \qquad \boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\mathbf{T}^{-1} \qquad \boldsymbol{\Phi}^* = \mathbf{T}\boldsymbol{\Phi}\mathbf{T}'$$

Then we have identically

$$\boldsymbol{\Lambda}^*\boldsymbol{\xi}^* \equiv \boldsymbol{\Lambda}\boldsymbol{\xi} \qquad \boldsymbol{\Lambda}^*\boldsymbol{\Phi}^*\boldsymbol{\Lambda}^{*'} \equiv \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'$$

This shows that at least k^2 independent conditions must be imposed on $\boldsymbol{\Lambda}$ and/or $\boldsymbol{\Phi}$ to make these identified.

Karl G Jöreskog Factor Analysis and Its Extensions

Slide 15

Two-Stage Least-Squares

$$y = \boldsymbol{\gamma}'\mathbf{x} + u , \tag{19}$$

$$\hat{\boldsymbol{\gamma}} = \mathbf{S}_{xx}^{-1}\mathbf{s}_{xy} , \tag{20}$$

$$\hat{\boldsymbol{\gamma}} = (\mathbf{S}'_{zx}\mathbf{S}_{zz}^{-1}\mathbf{S}_{zx})^{-1}\mathbf{S}'_{zx}\mathbf{S}_{zz}^{-1}\mathbf{s}_{zy} , \tag{21}$$

$$(n-p)^{-1}\hat{\sigma}_{uu}(\mathbf{S}'_{zx}\mathbf{S}_{zz}^{-1}\mathbf{S}_{zx})^{-1} , \tag{22}$$

$$\hat{\sigma}_{uu} = s_{yy} - 2\hat{\boldsymbol{\gamma}}'\mathbf{s}_{xy} + \hat{\boldsymbol{\gamma}}'\mathbf{S}_{xx}\hat{\boldsymbol{\gamma}} \tag{23}$$

Slide 16

Reference Variables Solution

Partitioning $\mathbf{x}$ into two parts $\mathbf{x}_1(k \times 1)$ and $\mathbf{x}_2(q \times 1)$, where $q = p - k$, and $\boldsymbol{\delta}$ similarly into $\boldsymbol{\delta}_1(k \times 1)$ and $\boldsymbol{\delta}_2(q \times 1)$, (17) can be written

$$\mathbf{x}_1 = \boldsymbol{\xi} + \boldsymbol{\delta}_1 \tag{24}$$

$$\mathbf{x}_2 = \boldsymbol{\Lambda}_2\boldsymbol{\xi} + \boldsymbol{\delta}_2 , \tag{25}$$

where $\boldsymbol{\Lambda}_2(q \times k)$ consists of the last $q = p - k$ rows of $\boldsymbol{\Lambda}$. The matrix $\boldsymbol{\Lambda}_2$ may, but need not, contain *a priori* specified elements. We say that the model is *unrestricted* when $\boldsymbol{\Lambda}_2$ is entirely unspecified and that the model is *restricted* when $\boldsymbol{\Lambda}_2$ contains *a priori* specified elements.

Slide 17

Solving (24) for $\boldsymbol{\xi}$ and substituting this into (25) gives

$$\mathbf{x}_2 = \boldsymbol{\Lambda}_2\mathbf{x}_1 + \mathbf{u}\ , \tag{26}$$

where $\mathbf{u} = \boldsymbol{\delta}_2 - \boldsymbol{\Lambda}_2\boldsymbol{\delta}_1$. Each equation in (26) is of the form (19) but it is not a regression equation because $\mathbf{u}$ is correlated with $\mathbf{x}_1$, since $\boldsymbol{\delta}_1$ is correlated with $\mathbf{x}_1$.

Let

$$x_i = \lambda_i^{'}\mathbf{x}_1 + u_i\ , \tag{27}$$

be the i-th equation in (26), where $\lambda_i^{'}$ is the i-th row of $\boldsymbol{\Lambda}_2$, and let $\mathbf{x}_{(i)}(q-1\times 1)$ be a vector of the remaining variables in $\mathbf{x}_2$. Then u_i is uncorrelated with $\mathbf{x}_{(i)}$ so that $\mathbf{x}_{(i)}$ can be used as instrumental variables for estimating (27). Provided $q \geq k+1$, this can be done for each $i = 1, 2, \ldots, q$.

Slide 18

Confirmatory Factor Analysis

In a confirmatory factor analysis, the investigator has such knowledge about the factorial nature of the variables that he/she is able to specify that each measure x_i depends only on a few of the factors ξ_j. If x_i does not depend on ξ_j, $\lambda_{ij} = 0$ in (16) (Slide 12). In many applications, the latent factor ξ_j represents a theoretical construct and the observed measures x_i are designed to be indicators of this construct. In this case there is only one non-zero λ_{ij} in each equation (16).

Karl G Jöreskog Factor Analysis and Its Extensions

Slide 19

Multigroup Analysis

Consider data from several groups or populations. These may be different nations, states, or regions, culturally or socioeconomically different groups, groups of individuals selected on the basis of some known selection variables, groups receiving different treatments, and control groups, etc. In fact, they may be any set of mutually exclusive groups of individuals that are clearly defined. It is assumed that a number of variables have been measured on a number of individuals from each population. This approach is particularly useful in comparing a number of treatment and control groups regardless of whether individuals have been assigned to the groups randomly or not.

Slide 20

Factorial Invariance

Consider the situation where the same tests have been administered in G different groups and the factor analysis model applied in each group:

$$\mathbf{x}_g = \boldsymbol{\Lambda}_g\boldsymbol{\xi}_g + \boldsymbol{\delta}_g,\ g = 1, 2, \ldots, G \tag{28}$$

The covariance matrix in group g is

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g\boldsymbol{\Phi}_g\boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g^2 \tag{29}$$

Hypothesis of factorial invariance:

$$\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 = \cdots = \boldsymbol{\Lambda}_G \tag{30}$$

Slide 21

Factorial Invariance with Latent Means

Sörbom (1974) extended the model in Slide 20 to include intercepts $\boldsymbol{\tau}$ in (28):

$$\mathbf{x}_g = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\xi}_g + \boldsymbol{\delta}_g,\ g = 1, 2, \ldots, G \tag{31}$$

Under complete factorial invariance:

$$\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_G \tag{32}$$

$$\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 = \cdots = \boldsymbol{\Lambda}_G \tag{33}$$

he showed that one can estimate the mean vector and covariance matrix of $\boldsymbol{\xi}$ in each group on a scale common to all groups.

Slide 22

Let $\bar{\mathbf{z}}_g$ and $\mathbf{S}_g$ be the sample mean vector and covariance matrix in group g, and let $\boldsymbol{\mu}_g(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_g(\boldsymbol{\theta})$ be the corresponding population mean vector and covariance matrix $g = 1, 2, \ldots, G$. The fit function for the multigroup case is defined as

$$\mathsf{F}(\boldsymbol{\theta}) = \sum_{g=1}^{G} \frac{N_g}{N} \mathsf{F}_g(\boldsymbol{\theta}) \ , \tag{34}$$

where $\mathsf{F}_g(\boldsymbol{\theta}) = \mathsf{F}(\bar{\mathbf{z}}_g, \mathbf{S}_g, \boldsymbol{\mu}_g(\boldsymbol{\theta}), \boldsymbol{\Sigma}_g(\boldsymbol{\theta}))$ is any of the fit functions defined for a single group. Here N_g is the sample size in group g and $N = N_1 + N_2 + \ldots + N_G$ is the total sample size. To test the model, one can again use $c = (N-1)$ times the minimum of F as a χ^2 with degrees of freedom $d = Gk(k+1)/2 - t$, where k is the number of variables.

Karl G Jöreskog Factor Analysis and Its Extensions

Slide 23

Econometric Models

$$\mathbf{y}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y}_t + \boldsymbol{\Gamma}\mathbf{x}_t + \mathbf{z}_t \tag{35}$$

$$t = 1, 2, \ldots, N \tag{36}$$

$$y_{ti} = \alpha_i + \boldsymbol{\beta}_{(i)} \mathbf{y}_{t(i)} + \boldsymbol{\gamma}_{(i)} \mathbf{x}_{t(i)} + z_{ti} \ , \tag{37}$$

$$\boldsymbol{\Sigma} = Cov \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}'\mathbf{A}' + \mathbf{A}\boldsymbol{\Psi}\mathbf{A}' & \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}'\mathbf{A}' & \boldsymbol{\Phi} \end{pmatrix} \tag{38}$$

Slide 24

Some History of LISREL

The idea of combining features of both *econometrics* and *psychometrics* into a single mathematical model was born in my mind in the spring of 1970. This idea was inspired by work of Professor Arthur S. Goldberger published in *Psychometrika*, 1971. The first version of LISREL was a *linear structural equation model* for *latent variables,* each with a single observed, possibly *fallible*, *indicator*. I presented this model at the conference on *Structural Equation Models in the Social Sciences* held in Madison, Wisconsin, in November 1970. The proceedings of this conference, edited by Professors Goldberger and Duncan, were published in 1973.

Slide 25

This LISREL model was generalized in 1971-72 to include models previously developed for *multiple indicators of latent variables*, for *confirmatory factor analysis*, for *simultaneous factor analysis in several populations* and more general models for *covariance structures*. The basic form of the LISREL model has remained the same ever since and is still the same model as used today. The general form of the LISREL model, due to its flexible specification in terms of fixed and free parameters and simple equality constraints, has proven to be so rich that it can handle not only the large variety of problems studied by hundreds of behavioral science researchers but also complex models, such as multiplicative MTMM models, non-linear models, and time series models, far beyond the type of models for which it was originally conceived.

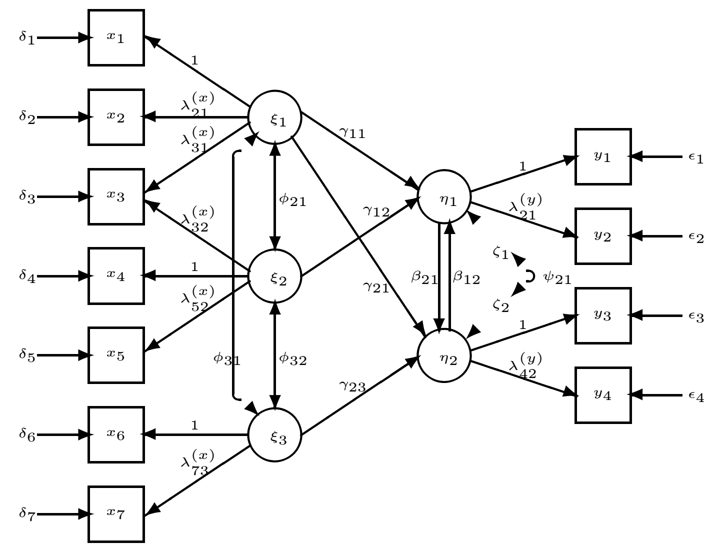
Slide 26

The first version of LISREL made generally available and with a written manual was LISREL III. It had fixed column input, fixed dimensions, only the maximum likelihood method, and users had to provide starting values for all parameters. The versions that followed demonstrated an enormous development in both statistical methodology and programming technology:

- LISREL IV (1978) had Keywords, Free Form Input, and Dynamic Storage Allocation
- LISREL V (1981) had Automatic Starting Values, Unweighted and Generalized Least Squares, and Total Effects
- LISREL VI (1984) had Parameter Plots, Modification Indices, and Automatic Model Modification
- LISREL 7 (1988) had PRELIS, Weighted Least Squares, and Completely Standardized Solution
- LISREL 8 (1994) had SIMPLIS, Path Diagrams, and Non-linear Constraints

Karl G Jöreskog Factor Analysis and Its Extensions

Slide 27



The LISREL Model in LISREL Notation

Slide 28

The LISREL Model with Means

$$\mathbf{y} = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\epsilon}$$

$$\mathbf{x} = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}$$

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

$\mathbf{y}, \mathbf{x}$ = Observed Variables $\boldsymbol{\eta}, \boldsymbol{\xi}$ = Latent Variables

$\boldsymbol{\epsilon}, \boldsymbol{\delta}$ = Measurement Errors $\boldsymbol{\zeta}$ = Structural Errors

$\boldsymbol{\tau}_y, \boldsymbol{\tau}_x, \boldsymbol{\alpha}$ = Intercept Terms

$\boldsymbol{\Lambda}_y, \boldsymbol{\Lambda}_x, \mathbf{B}, \boldsymbol{\Gamma}$ = Parameter Matrices

Slide 29

Assumptions

- $\boldsymbol{\epsilon}$ is uncorrelated with $\boldsymbol{\eta}$
- $\boldsymbol{\delta}$ is uncorrelated with $\boldsymbol{\xi}$
- $\boldsymbol{\zeta}$ is uncorrelated with $\boldsymbol{\xi}$
- $\boldsymbol{\zeta}$ is uncorrelated with $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$
- $\mathbf{I} - \mathbf{B}$ is non-singular

Slide 30

LISREL as a Mean and Covariance Structure

Let

$$\boldsymbol{\kappa} = E(\boldsymbol{\xi}) \quad \boldsymbol{\Phi} = Cov(\boldsymbol{\xi})$$

$$\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\Theta}_{\epsilon} & \boldsymbol{\Theta}'_{\delta\epsilon} \\ \boldsymbol{\Theta}_{\delta\epsilon} & \boldsymbol{\Theta}_{\delta} \end{pmatrix} = Cov \begin{pmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\delta} \end{pmatrix}$$

Then it follows that the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{z} = (\mathbf{y}', \mathbf{x}')'$ are

Karl G Jöreskog Factor Analysis and Its Extensions

Slide 31

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\kappa}) \\ \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x\boldsymbol{\kappa} \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Lambda}_y\mathbf{A}(\boldsymbol{\Gamma\Phi\Gamma}' + \boldsymbol{\Psi})\mathbf{A}'\boldsymbol{\Lambda}'_y + \boldsymbol{\Theta}_{\epsilon} & \boldsymbol{\Lambda}_y\mathbf{A}\boldsymbol{\Gamma\Phi\Lambda}'_x + \boldsymbol{\Theta}'_{\delta\epsilon} \\ \boldsymbol{\Lambda}_x\boldsymbol{\Phi\Gamma}'\mathbf{A}'\boldsymbol{\Lambda}'_y + \boldsymbol{\Theta}_{\delta\epsilon} & \boldsymbol{\Lambda}_x\boldsymbol{\Phi\Lambda}'_x + \boldsymbol{\Theta}_{\delta} \end{pmatrix},$$

where $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$.

The elements of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are functions of the elements of $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}$, $\boldsymbol{\tau}_y$, $\boldsymbol{\tau}_x$, $\boldsymbol{\Lambda}_y$, $\boldsymbol{\Lambda}_x$, $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Theta}_{\epsilon}$, $\boldsymbol{\Theta}_{\delta}$, and $\boldsymbol{\Theta}_{\delta\epsilon}$ which are of three kinds:

- *fixed parameters* that have been assigned specified values,
- *constrained parameters* that are unknown but linear or non-linear functions of one or more other parameters, and
- *free parameters* that are unknown and not constrained.

Slide 32

General Mean and Covariance Structures

Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be functions of a parameter vector $\boldsymbol{\theta}$:

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}) \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) \tag{39}$$

or

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}) \quad \boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta}) \tag{40}$$

where $\boldsymbol{\sigma}$ is a vector of the non-duplicated elements of $\boldsymbol{\Sigma}$.

Slide 33

Some Formulas

Let **s** be a vector of the non-duplicated elements of **S** and assume that

$$n^{\frac{1}{2}}(\mathbf{s}-\boldsymbol{\sigma}) \rightarrow N(\mathbf{0},\boldsymbol{\Omega}) \quad (41)$$

Definitions

$$\begin{aligned} k &= \text{number of observed variables} \\ s &= \frac{1}{2}k(k+1) \\ t &= \text{number of independent parameters} < s \\ d &= s-t \end{aligned}$$

Slide 34

$$\underset{k^2\times s}{\mathbf{K}} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\text{, where } \underset{k^2\times s}{\mathbf{D}} \text{ is the duplication matrix}$$

$$\underset{s\times 1}{\mathbf{s}} = \mathbf{K}'\mathsf{vec}(\mathbf{S}) \qquad \mathsf{vec}(\mathbf{S}) = \mathbf{Ds}$$

$$\underset{s\times s}{\boldsymbol{\Omega}} = n\mathsf{ACov}(\mathbf{s}) \qquad n = N-1$$

$$\underset{s\times s}{\mathbf{W}} = n\,\mathsf{Est}[\mathsf{ACov}(\mathbf{s})]$$

$$\begin{aligned} \mathbf{W}_{\mathsf{NT}} &= 2\mathbf{K}'(\hat{\boldsymbol{\Sigma}}\otimes\hat{\boldsymbol{\Sigma}})\mathbf{K} \text{ under } \mathsf{NT}, \\ \mathbf{W}_{\mathsf{NNT}} &= (w_{gh,ij}) \text{ under } \mathsf{NNT}, \\ w_{gh,ij} &= n\mathsf{Est}[\mathsf{ACov}(s_{gh},s_{ij})] = m_{ghij} - s_{gh}s_{ij}\ , \\ m_{ghij} &= (1/N)\sum_{a=1}^{N}(z_{ag}-\bar{z}_g)(z_{ah}-\bar{z}_h)(z_{ai}-\bar{z}_i)(z_{aj}-\bar{z}_j) \end{aligned}$$

Karl G Jöreskog Factor Analysis and Its Extensions

Slide 35

$$\boldsymbol{\sigma} = \underset{s\times 1}{\boldsymbol{\sigma}}\ (\underset{t\times 1}{\boldsymbol{\theta}})$$

$$\underset{s\times t}{\boldsymbol{\Delta}} = \frac{\partial\boldsymbol{\sigma}}{\partial\boldsymbol{\theta}'} \text{ evaluated at } \hat{\boldsymbol{\theta}}$$

$$\underset{s\times d}{\boldsymbol{\Delta}_c} = \text{orthogonal complement to } \boldsymbol{\Delta}$$

$$\boldsymbol{\Delta}_c'\boldsymbol{\Delta} = \mathbf{0} \qquad [\boldsymbol{\Delta}|\boldsymbol{\Delta}_c] \quad \text{non-singular}$$

Slide 36

Fit Functions

$$\mathsf{F} = (\mathbf{s}-\boldsymbol{\sigma})'\underset{s\times s}{\mathbf{V}}(\mathbf{s}-\boldsymbol{\sigma})\ ,$$

where the weight matrix **V** is defined differently for different fit functions:

ULS: $\mathbf{V} = \mathbf{I}^* = \mathsf{diag}\ (1,2,1,2,2,1,\ldots)$

GLS: $\mathbf{V} = \mathbf{D}'(\mathbf{S}^{-1}\otimes\mathbf{S}^{-1})\mathbf{D}$

ML: $\mathbf{V} = \mathbf{D}'(\hat{\boldsymbol{\Sigma}}^{-1}\otimes\hat{\boldsymbol{\Sigma}}^{-1})\mathbf{D}$

WLS: $\mathbf{V} = \mathbf{W}_{\mathsf{NNT}}^{-1}$ or $\mathbf{W}_{\mathsf{NNT}}^{-}$ if $\mathbf{W}_{\mathsf{NNT}}$ singular

DWLS: $\mathbf{V} = \mathbf{D}_{\mathsf{W}}^{-1} = [\mathsf{diag}\mathbf{W}]^{-1}$

with $\mathbf{W} = \mathbf{W}_{\mathsf{NT}}$ or

$\mathbf{W} = \mathbf{W}_{\mathsf{NNT}}$

Slide 37

Results

$$\underset{s\times s}{\mathbf{E}} = \mathbf{\Delta}'\mathbf{V}\mathbf{\Delta}$$

$$n\mathsf{ACov}(\hat{\boldsymbol{\theta}}) = \mathbf{E}^{-1}\mathbf{\Delta}'\mathbf{V}\mathbf{W}\mathbf{V}\mathbf{\Delta}\mathbf{E}^{-1}$$

$$n\mathsf{ACov}(\mathbf{s}-\hat{\boldsymbol{\sigma}}) = \mathbf{W} - \mathbf{\Delta}\mathbf{E}^{-1}\mathbf{\Delta}'$$

with $\mathbf{W} = \mathbf{W}_{\mathsf{NT}}$ or $\mathbf{W} = \mathbf{W}_{\mathsf{NNT}}$

$$c_2 = n(\mathbf{s}-\hat{\boldsymbol{\sigma}})'\mathbf{\Delta}_c(\mathbf{\Delta}_c'\mathbf{W}_{\mathsf{NT}}\mathbf{\Delta}_c)^{-1}\mathbf{\Delta}_c'(\mathbf{s}-\hat{\boldsymbol{\sigma}})$$

$$h_1 = \mathsf{tr}[(\mathbf{\Delta}_c'\mathbf{W}_{\mathsf{NT}}\mathbf{\Delta}_c)^{-1}(\mathbf{\Delta}_c'\mathbf{W}_{\mathsf{NNT}}\mathbf{\Delta}_c)]$$

$$c_3 = (d/h_1)c_2$$

$$c_4 = n(\mathbf{s}-\hat{\boldsymbol{\sigma}})'\mathbf{\Delta}_c(\mathbf{\Delta}_c'\mathbf{W}_{\mathsf{NNT}}\mathbf{\Delta}_c)^{-1}\mathbf{\Delta}_c'(\mathbf{s}-\hat{\boldsymbol{\sigma}})$$

Slide 38

The Success of Structural Equation Modeling

There has been an enormous development of structural equation modeling in the last 30 years.

Proof:

- Thousands of journal articles
- Hundreds of dissertations
- Numerous books

Karl G Jöreskog Factor Analysis and Its Extensions

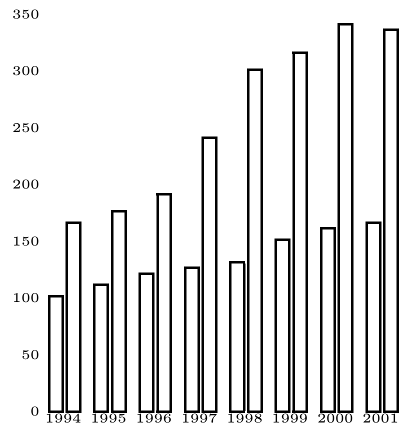
Slide 39

What are the factors behind this development?

- Models can be tested
- Computer technology
- Simple command language
- Path diagram
- SEM courses at many universities
- Journal of Structural Equation Modeling

Slide 40

The Growth of Structural Equation Modeling



Number of Journals and Articles by Year

Slide 41

Main Virtues of SEM Methodology

- SEM has the power to test complex hypotheses involving causal relationships among construct or latent variables
- SEM unifies several multivariate methods into one analytic framework
- SEM specifically expresses the effects of latent variables on each other and the effect of latent variables on observed variables
- SEM can be used to test alternative hypotheses.

SEM gives social and behavioral researchers powerful tools for

- stating theories more exactly,
- testing theories more precisely,
- generating a more thorough understanding of observed data.

Slide 42

Retraction

Is it really that great?

In the preface of his 1975 book O. D. Duncan said that he "was fascinated by the formal properties of causal models but held a rather agnostic view of their utility".

We have certainly come to great strides in the "formal" realm but are there really any great substantive applications?

Karl G Jöreskog Factor Analysis and Its Extensions

Slide 43

Closing the Circle

Back to Factor Analysis

Latent Variable Models

$$f(\mathbf{x}) = \int h(\boldsymbol{\xi}) g(\mathbf{x} \mid \boldsymbol{\xi}) d\boldsymbol{\xi} \tag{42}$$

$$f(\mathbf{x}) = \int h(\boldsymbol{\xi}) \prod_{i=1}^{p} g(x_i \mid \boldsymbol{\xi}) d\boldsymbol{\xi} \tag{43}$$

$$\boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{I}) \quad \mathbf{x} \mid \boldsymbol{\xi} \sim N(\boldsymbol{\mu} + \boldsymbol{\Lambda\xi}, \boldsymbol{\Psi}^2) \Rightarrow \mathbf{x} = N(\boldsymbol{\mu}, \boldsymbol{\Lambda\Lambda}' + \boldsymbol{\Psi}^2) \tag{44}$$

Slide 44

Binary and Ordinal Variables

$$x_i = 1, 2, \ldots m_i \tag{45}$$

$$g_i(x_i = s \mid \boldsymbol{\xi}) = F(\alpha_s^{(i)} - \sum_{j=1}^{k} \beta_{ij}\xi_j) - F(\alpha_{s-1}^{(i)} - \sum_{j=1}^{k} \beta_{ij}\xi_j) \tag{46}$$

$$-\infty = \alpha_0^{(i)} < \alpha_1^{(i)} < \alpha_2^{(i)} \cdots < \alpha_{m_{i-1}}^{(i)} < \alpha_{m_i}^{(i)} = \infty$$

$$\text{NOR}: \ F(t) = \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \tag{47}$$

$$\text{POM}: \ F(t) = \Psi(t) = \frac{e^t}{1 + e^t} \tag{48}$$