50 Years of SEM in 50 Minutes??

Karl G Jöreskog

Norwegian Business School & Uppsala University

May 11, 2015

Factor Analysis before 1964

- Although its roots can be traced back to the work of Francis Galton, it is generally considered that factor analysis began with the celebrated article by Spearman (1904). In the first half of the 20th century factor analysis was mainly developed by psychologists for the purpose of identifying mental abilities by means of psychological testing.
- Various theories of mental abilities and various procedures for analyzing the correlations among psychological tests emerged. The most prominent factor analysts in the first half of the 20th century seem to be Godfrey Thomson, Cyril Burt, Raymond Cattell, Karl Holzinger, Louis Thurstone and Louis Guttman. A later generation of psychological factor analysts that played important roles are Ledyard Tucker, Ray Cattell, Henry Kaiser, and Chester Harris.

- Through the 1950's factor analysis was characterized by a set of *ad hoc* procedures for analyzing the correlation matrix **R** of the tests.
- Four problems of factor analysis emerged:
  - Number of factors
  - Communalities
  - Factor extraction
  - Factor rotation
- The focus was om computation. Computers were very rare and consisted of large mainframes that filled whole rooms.

- A few statisticians had began to be interested in factor analysis, notably
- Lawley, D.N. (1940) The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society Edinburgh*, **60**, 64–82.
- Anderson, T.W., and Rubin, H. (1956) Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium*, Volume V. Berkeley: University of California Press.
- Jöreskog, K.G. (1962) On the statistical treatment of residuals in factor analysis. *Psychometrika, 27,* 335-345.
- Jöreskog, K.G. (1963) *Statistical Estimation in Factor Analysis: A New Technique and its Foundation*. Stockholm: Almqvist & Wiksell.

Communalities

Guttman (1953) defined the factor analysis problem as follows. What numbers should be put in the diagonal of $\mathbf{R}$ such that this matrix is Gramian and of smallest possible rank k. This is equivalent to finding a matrix $\mathbf{\Lambda}$ of order $p \times k$, where $k < p$, such that

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

where $\mathbf{R}_c$ is $\mathbf{R}$ with communalities in the diagonal.

The problem of communalities was involved in much discussion of factor analysis in the 1950's.

Factor Extraction

Once the communalities have been determined, one could determine $\mathbf{\Lambda}$ in (1). The most common method in the early literature is one which chooses the columns of $\mathbf{\Lambda}$ proportional to the eigenvectors of $\mathbf{R}_c$ corresponding to the k largest eigenvalues.

After $\mathbf{\Lambda}$ has been determined, the communalities can be re-estimated as the sum of squares of each row in $\mathbf{\Lambda}$. Putting these new communalities in the diagonal of $\mathbf{R}$ gives a new matrix $\mathbf{R}_c$ from which a new $\mathbf{\Lambda}$ can be obtained. This process can be repeated. In this process it can happen that one or more of the communalities exceed 1, so called *Heywood cases*. Such Heywood cases occurred quite often in practice and caused considerable problems.

The Factor Analysis Model

The basic idea of factor analysis is the following. For a given set of observed response variables $x_1, \ldots, x_p$ one wants to find a set of underlying latent factors $\xi_1, \ldots, \xi_k$, much fewer than the observed variables. These factors are supposed to account for the correlations of the response variables. This leads to the linear factor analysis model of Thurstone (1947):

$$x_i = \mu_i + \lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \cdots + \lambda_{ik}\xi_k + \delta_i \ , i = 1, 2, \ldots, p \ , \qquad (2)$$

where δ_i, the unique part of x_i, is uncorrelated with $\xi_1, \xi_2, \ldots, \xi_k$ and with δ_j for $j \neq i$. In matrix notation (2) is

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta} \ , \quad \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \ . \qquad (3)$$

The objective of factor analysis is to estimate the number of factors k and the factor loadings $\boldsymbol{\Lambda} = (\lambda_{ij})$ from a random sample of observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$.

Guttman's (1953) Image Theory

Guttman (1953) considered a different system than (2), namely the regression of x_i on all the other x's:

$$x_i = \mu_i + \beta_{i1} x_1 + \beta_{i2} x_2 + \cdots + \beta_{i,i-1} x_{i-1} + \beta_{i,i+1} x_{i+1} + \cdots + \beta_p x_p + z_i \ , \quad (4)$$

that is

$$x_i = \mu_i + \boldsymbol{\beta}'_{)i(} \mathbf{x}_{)i(} + z_i \ , \quad (5)$$

or in matrix form

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{B}\mathbf{x} + \mathbf{z} \ , \quad (6)$$

where $\mathbf{B}$ is a matrix of order $p \times p$ with $\beta_{ii} = 0$.
What does (4) has to do with (2)?
Guttman (1956) showed that the squared multiple correlation in the regression (4) is a lower bound for communality. Let's consider some statements which are equivalent to this.

For this purpose consider some notation

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{ii} & \\ \boldsymbol{\sigma}_i & \boldsymbol{\Sigma}_{ii} \end{pmatrix} \quad \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma^{ii} & \\ \boldsymbol{\sigma}^i & \boldsymbol{\Sigma}^{ii} \end{pmatrix} \tag{7}$$

$$\sigma^{ii} = (\sigma_{ii} - \boldsymbol{\sigma}_i' \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\sigma}_i)^{-1} \tag{8}$$

$$i = 1, 2, \ldots, p$$

$$x_i = c_i + \delta_i, \;\; \delta_i \perp c_i, \;\; \delta_i \perp \delta_j \; j \neq i \tag{9}$$

$$x_i = p_i + z_i, \;\; z_i \perp p_i \tag{10}$$

$$\sigma_{ii} = Var(c_i) + Var(\delta_i) \tag{11}$$

$$\sigma_{ii} = Var(p_i) + Var(z_i) \tag{12}$$

$$R_i^2 = \frac{Var(p_i)}{\sigma_{ii}} \leq \frac{Var(c_i)}{\sigma_{ii}} \Leftrightarrow Var(p_i) \leq Var(c_i) \Leftrightarrow Var(\delta_i) \leq Var(z_i) \tag{13}$$

But

$$Var(z_i) = \sigma_{ii} - \boldsymbol{\sigma}_i' \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\sigma}_i = 1/\sigma^{ii} \tag{14}$$

so

$$\psi_i \leq 1/\sigma^{ii} \Leftrightarrow \psi_{ii}\sigma^{ii} \leq 1 \tag{15}$$

This leads to the model

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \theta(diag\boldsymbol{\Sigma}^{-1})^{-1} \, , \tag{16}$$

which is to be interpreted as an implicit equation defining $\boldsymbol{\Sigma}$ as a function of $\boldsymbol{\Lambda}$ and θ. In my dissertation I developed a simple non-iterative method for estimating $\boldsymbol{\Lambda}$ and θ.

Pre- and postmultiplying (16) by $(diag\boldsymbol{\Sigma}^{-1})^{\frac{1}{2}}$ and defining

$$\boldsymbol{\Sigma}^{\star} = (diag\boldsymbol{\Sigma}^{-1})^{\frac{1}{2}}\boldsymbol{\Sigma}(diag\boldsymbol{\Sigma}^{-1})^{\frac{1}{2}} \, ,$$

and

$$\boldsymbol{\Lambda}^{\star} = (diag\boldsymbol{\Sigma}^{-1})^{\frac{1}{2}}\boldsymbol{\Lambda}$$

gives

$$\boldsymbol{\Sigma}^{\star} = \boldsymbol{\Lambda}^{\star}\boldsymbol{\Lambda}^{\star\prime} + \theta\mathbf{I} \, ,$$

which shows that $p - k$ of the eigenvalues of $\boldsymbol{\Sigma}^{\star}$ are equal to θ.

Let $\mathbf{S}$ be a consistent estimate of $\mathbf{\Sigma}$. Then

$$\mathbf{S}^{\star} = (diag\mathbf{S}^{-1})^{\frac{1}{2}}\mathbf{S}(diag\mathbf{S}^{-1})^{\frac{1}{2}} ,$$

is a consistent estimate of $\mathbf{\Sigma}^{\star}$. Let $\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_p$ be the eigenvalues of $\mathbf{S}^{\star}$ in descending order and let $\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_2, \ldots, \hat{\boldsymbol{\omega}}_k$ be unit-length eigenvectors corresponding to the k largest eigenvalues. Furthermore, let

$$\hat{\mathbf{\Gamma}}_k = diag(\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_k) ,$$

and

$$\hat{\mathbf{\Omega}}_k = (\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_2, \ldots, \hat{\boldsymbol{\omega}}_k) .$$

Then the simple solution is

$$\hat{\theta} = \frac{1}{p-k}(\hat{\gamma}_{k+1} + \hat{\gamma}_{k+2} + \cdots + \hat{\gamma}_p) ,$$

$$\hat{\mathbf{\Lambda}} = (diag\mathbf{S}^{-1})^{-\frac{1}{2}}\hat{\mathbf{\Omega}}_k(\hat{\mathbf{\Gamma}}_k - \hat{\theta}\mathbf{I})^{\frac{1}{2}}\mathbf{U} ,$$

where $\mathbf{U}$ is an arbitrary orthogonal matrix of order $k \times k$. This solution also offers a solution to the number of factors problem. Choose the smallest k such that $\hat{\theta} < 1$.

This simple solution has several obvious advantages:

- It is non-iterative and very fast to compute.
- It does not require estimates of communalities.
- Heywood-cases cannot occur, *i.e.*, the estimates of uniquenesses which are the diagonal elements in $\hat{\theta} diag \mathbf{S}^{-1}$ are always positive.
- It is scale-free in the sense that if $\mathbf{x}$ is replaced by $\mathbf{Dx}$, where $\mathbf{D}$ is a diagonal matrix of scale factors, then $\hat{\mathbf{\Lambda}}$ will be replaced by $\mathbf{D}\hat{\mathbf{\Lambda}}$ while $\hat{\theta}$ is unchanged.

Note that the matrix $\mathbf{S}^{\star}$ is independent of $\mathbf{D}$, yet it is not a correlation matrix. The part $\hat{\mathbf{\Omega}}_k(\hat{\mathbf{\Gamma}}_k - \hat{\theta}\mathbf{I})^{\frac{1}{2}}\mathbf{U}$ of the solution is also independent of $\mathbf{D}$.

Later I also developed a maximum likelihood method for this model. see Jöreskog, K.G. (1969) Efficient estimation in image factor analysis. *Psychometrika*, **34**, 51–75. But this went unnoticed, why?

Maximum Likelihood Factor Analysis

Jöreskog, K.G. (1967) Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32**, 443–482.

*****

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, be iid with $\mathbf{x}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ positive definite. If $\boldsymbol{\mu}$ is unconstrained and

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \ . \tag{17}$$

then maximizing $\ln L$ is equivalent to minimizing

$$F(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) = \log \|\boldsymbol{\Sigma}\| + tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log \|\mathbf{S}\| - p \ , \tag{18}$$

Jöreskog (1967) approached the computational problem by focusing on the concentrated fit function

$$f(\mathbf{\Psi}) = \min_{\mathbf{\Lambda}} F(\mathbf{\Lambda}, \mathbf{\Psi}) \ , \tag{19}$$

which could be minimized numerically.

If one or more of the ψ_i gets close to zero, this procedure becomes unstable, a problem that can be circumvented by reparameterizing:

$$\theta_i = \ln \psi_i \ , \ \ \psi_i = +e^{\theta_i} \ . \tag{20}$$

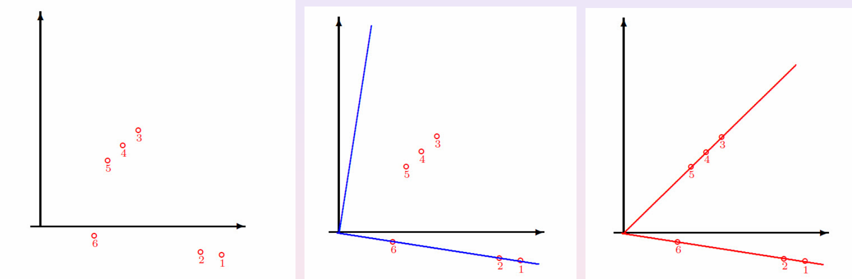
This leads to a very fast and efficient algorithm.

Rotation

When $k > 1$, the factor loadings in $\mathbf{\Lambda}$ are not uniquely defined. Geometrically the factor loadings may be viewed as p points in a k-dimensional space. In this space the points are fixed but their coordinates can be referred to different factor axes. If the factor axes are orthogonal we say we have an *orthogonal solution*; if they are oblique we say that we have an *oblique solution* where the cosine of the angles between the factor axes are interpreted as correlations between the factors.

In statistical terminology, an orthogonal solution corresponds to *uncorrelated factors* and an oblique solution corresponds to *correlated factors*. One can also have solutions in which some factors are uncorrelated and some are correlated.

Rotation is illustrated in the following figures



Factor Rotation: Unrotated Solution

Factor Rotation: Orthogonal Solution

Factor Rotation: Oblique Solution

To facilitate the interpretation of the factors one makes an orthogonal or oblique rotation of the factor axes. This rotation is usually guided by Thurstone's principle of simple structure which essentially states that only a small fraction of the loadings in each row and column should be large. Geometrically, this means that the factor axes pass through or near as many points as possible.

*****

Jöreskog, K.G. (1966) Testing a simple structure hypothesis in factor analysis. *Psychometrika*, **31**, 165-178.

In exploratory factor analysis it is usually assumed that the factors $\xi_1, \ldots, \xi_k$ are uncorrelated and have variances 1. These assumptions can be relaxed and the factors may be correlated and they need not have variance 1. If $\boldsymbol{\xi}$ has covariance matrix $\boldsymbol{\Phi}$, the covariance matrix of $\mathbf{x}$ is

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \, . \tag{21}$$

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

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

Since $\mathbf{T}$ has k^2 independent elements, this shows that at least k^2 independent conditions must be imposed on $\boldsymbol{\Lambda}$ and/or $\boldsymbol{\Phi}$ to make these identified.

Factor analysis was typically done in two steps. In the first step, one obtains an arbitrary orthogonal solution in which $\mathbf{\Phi} = \mathbf{I}$ in (21). In the second step, this is rotated orthogonally or obliquely to achieve a simple structure. For the rotated factors to have unit variance, $\mathbf{T}$ must satisfy

$$diag(\mathbf{T}\mathbf{T}') = \mathbf{I} \,, \tag{22}$$

for an oblique solution and

$$\mathbf{T}\mathbf{T}' = \mathbf{I} \,, \tag{23}$$

for an orthogonal solution.

Confirmatory Factor Analysis

Jöreskog, K.G. (1969) A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, **34**, 183-202.

*****

In contrast to exploratory factor analysis, a confirmatory factor analysis begins by defining the latent variables one would like to measure. This is based on substantive theory and/or previous knowledge. One then constructs observable variables to measure these latent variables. Thus, in a confirmatory factor analysis, the number of factors is known and equal to the number of latent variables. The confirmatory factor analysis is a model that should be estimated and tested.
Exploratory and confirmatory factor analysis are illustrated in Figures 1 and 2.

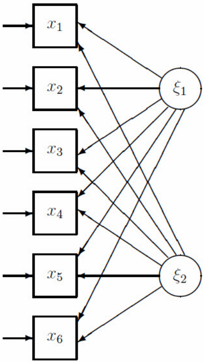


Figure 1: Exploratory Factor Analysis

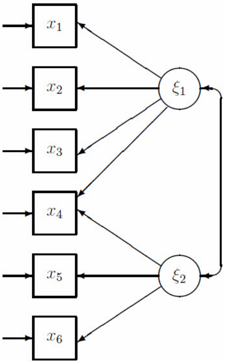


Figure 2: Confirmatory Factor Analysis

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 (2). In many applications, the latent variable 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 (2). In general, assuming that $\mathbf{\Phi}$ is a correlation matrix, one needs to specify at least $k - 1$ zero elements in each column of $\mathbf{\Lambda}$ but in a confirmatory factor analysis there are usually many more zeros in each column.

The possibility of *a priori* specified zero elements in $\mathbf{\Lambda}$ was mentioned in Anderson & Rubin (1956) and in Jöreskog & Lawley (1968), but the term confirmatory factor analysis was first used in Jöreskog (1969).

To estimate a confirmatory factor analysis model one can minimize any of the fit function (18) with respect to all free elements of $\mathbf{\Lambda}$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$. In most cases no analytic solution is available so the minimization must be done numerically. By contrast to exploratory factor analysis, no eigenvalues and eigenvectors are involved and the solution is obtained in one step. No factor rotation is needed.

In a way, confirmatory factor analysis shifts the focus from the problems of factor extraction and rotation to the problem of testing a specified model. With the ML method, the most common way of testing the model is to use N times the minimum value of the fit function F_{ML} as a χ^2 with degrees of freedom equal to $\frac{1}{2}p(p+1)$ minus the number of independent parameters in $\mathbf{\Lambda}$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$.

Covariance Structures

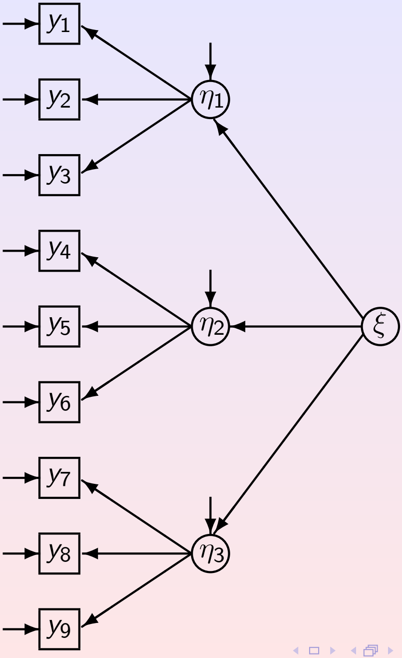
Jöreskog, K.G. (1974) Analyzing psychological data by structural analysis of covariance matrices. In R.C. Atkinson et al. (Eds.): *Contemporary Developments in Mathematical Psychology*- Volume II. San Francisco: W.H. Freeman, 1–56.

*****

Equation (21) can be extended in various ways, for example,

$$\mathbf{\Sigma} = \mathbf{\Lambda}_y(\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}' + \mathbf{\Psi})\mathbf{\Lambda}_y{}' + \mathbf{\Theta}_\epsilon \ . \tag{24}$$

This can accommodate second-order factor analysis, where $\mathbf{\Lambda}_y$ is the first-order factor loadings and $\mathbf{\Gamma}$ are the second-order factor loadings, see next slide.


y_1
y_2
y_3
y_4
y_5
y_6
y_7
y_8
y_9
η_1
η_2
η_3
ξ

Equation (21) can also accommodate various test theory models shown here

Model	Covariance Structure	No. of Parameters
Parallel	$\mathbf{\Sigma} = \lambda^2 \mathbf{jj}' + \theta \mathbf{I}$	2
Tau-equivalent	$\mathbf{\Sigma} = \lambda^2 \mathbf{jj}' + \mathbf{\Theta}$	$p + 1$
Variable-length	$\mathbf{\Sigma} = \mathbf{D}_\lambda(\lambda\lambda' + \psi \mathbf{I})\mathbf{D}_\lambda$	$p + 1$
Congeneric	$\mathbf{\Sigma} = \lambda\lambda' + \mathbf{\Theta}$	$2p$

Multi-Group Factor Analysis

Jöreskog, K.G. (1971) Simultaneous factor analysis in several populations. *Psychometrika*, 57, 409–426.

*****

Consider data from several groups or populations of individuals. 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.

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

$$\mathbf{x}_g = \boldsymbol{\Lambda}_g \boldsymbol{\xi}_g + \boldsymbol{\delta}_g, \; g = 1, 2, \ldots, G \; , \tag{25}$$

where, as before, $\boldsymbol{\xi}_g$ and $\boldsymbol{\delta}_g$ are uncorrelated. The covariance matrix of $\mathbf{x}_g$ in group g is

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g \boldsymbol{\Phi}_g \boldsymbol{\Lambda}'_g + \boldsymbol{\Psi}^2_g \; . \tag{26}$$

The hypothesis of factorial invariance is:

$$\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \cdots = \mathbf{\Lambda}_G \ . \tag{27}$$

This states that the factor loadings are the same in all groups. Group differences in variances and covariances of the observed variables are due only to differences in variances and covariances of the factors and different error variances. The idea of factorial invariance is that the factor loadings are attributes of the tests and they should therefore be independent of the population sampled, whereas the distribution of the factors themselves could differ across populations. A stronger assumption is to assume that the error variances are also equal across groups:

$$\mathbf{\Psi}_1 = \mathbf{\Psi}_2 = \cdots = \mathbf{\Psi}_G \ . \tag{28}$$

Structural Equation Models(SEM)

Jöreskog, K.G. (1973) A general method for estimating a linear structural equation system. In A.S. Goldberger and O.D. Duncan (Eds.): *Structural Equation Models in the Social Sciences*. New York: Seminar Press, 85–112.

*****

Factor analysis is used to investigate latent variables that are presumed to underlie a set of manifest variables. Understanding the structure and meaning of the latent variables in the context of their manifest variables is the main goal of traditional factor analysis. After a set of factors has been identified, it is natural to go on and use the factors themselves as predictors or outcome variables in further analyses. Broadly speaking, this is the goal of *structural equation modeling*.

A further extension of the classical factor analysis model is to allow the factors not only to be correlated, as in confirmatory factor analysis, but also to allow some latent variables to depend on other latent variables. Models of this kind are called structural equation models and there are many examples of this in the literature.

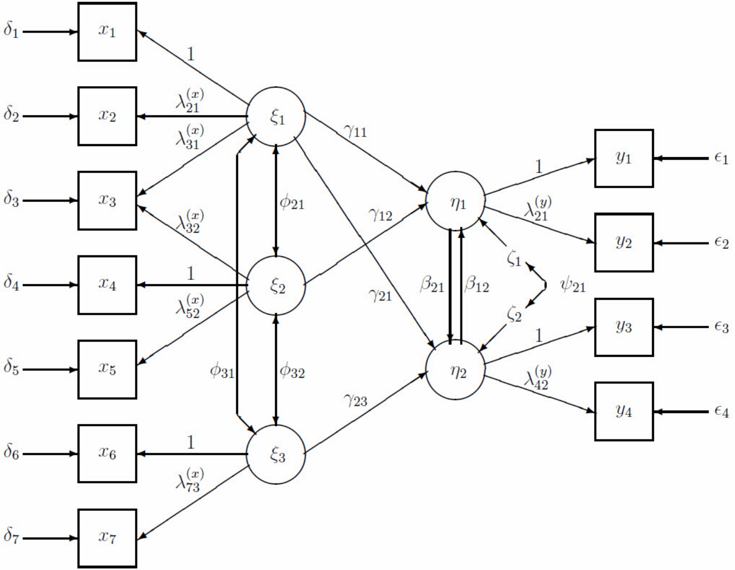


Figure 5: The LISREL Model

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}\ , \tag{29}$$

$$\mathbf{y} = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}\ , \tag{30}$$

and

$$\mathbf{x} = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}\ , \tag{31}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\kappa}) \\ \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x\boldsymbol{\kappa} \end{pmatrix}\ , \tag{32}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Lambda}_y\mathbf{B}^\star(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})\mathbf{B}^{\star\prime}\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon & \boldsymbol{\Lambda}_y\mathbf{B}^\star\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_{\delta\epsilon}' \\ \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Gamma}'\mathbf{B}^{\star\prime}\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_{\delta\epsilon} & \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta \end{pmatrix}\ , \tag{33}$$

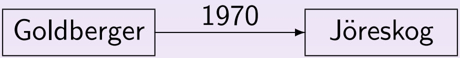
$$\mathbf{B}^\star = (\mathbf{I} - \mathbf{B})^{-1}$$

.

- *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.

The LISREL model combines features of both econometrics and psychometrics into a single model. The first LISREL model was a linear structural equation model for latent variables, each with a single observed, possibly fallible, indicator, see Jöreskog (1973). This model was generalized to models with multiple indicators of latent variables, to simultaneous structural equation models in several groups, and to more general covariance structures. Jöreskog & Sörbom developed the LISREL program.

Some History of LISREL



The idea of combining features of both *econometrics* and *psychometrics* into a single mathematical model was born in February 1970.

The first version of LISREL was a *linear structural equation model* for *latent variables,* each with a single observed, possibly *fallible*, *indicator*. This model was presented 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. This LISREL model was generalized in 1971-72 to include models previously developed for *multiple indicators of latent variables*

The basic form of the LISREL model has remained the same ever since and is still the same model as used today.

The first two computer versions of LISREL were written in 1970–71. The program was completely rewritten in 1974-75 by Dag Sörbom. This version, called LISREL III, was the first made generally available with a written manual. 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:

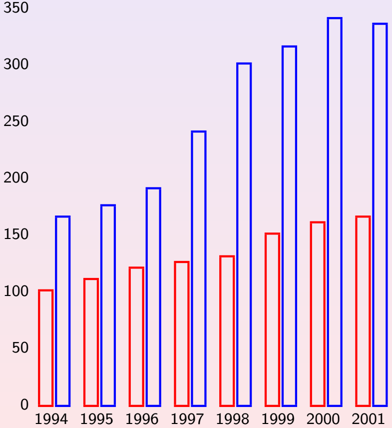
- LISREL IV (1978) had Keywords, Free Form Input, and Dynamic Storage Allocation
- LISREL V (1981) had Automatic Starting Values, Unweighted and Generalized Least Squares, and Total Effects
- LISREL VI (1984) had Parameter Plots, Modification Indices, and Automatic Model Modification
- LISREL 7 (1988) had PRELIS, Weighted Least Squares, and Completely Standardized Solution
- LISREL 8 (1994) had SIMPLIS, Path Diagrams, and Non-linear Constraints
- LISREL 9 (2013) with FIML for Missing Values and Adaptive Quadrature for Ordinal Variables

Two More Recent Important Developments

- Robust Estimation (Browne, 1984, Satorra & Bentler, 1988)
- Ordinal Variables
  - Underlying variables approach (Muthen ,1984, Jöreskog, 1990,1994)
  - Latent trait models (Jöreskog & Moustaki, 2001) estimated with adaptive quadrature (Schilling & Bock, 2005)

The Growth of Structural Equation Modeling

SEM became became very popular in multivariate analysis much because of the LISREL program. As witnessed by the literature, there has been an enormous development of both the statistical theory and computer technology, Hershberger (2003).



Number of Journals and Articles by Year

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.

Father of the LISREL Model turns 80

24.04.2015 Professor Karl Jøreskog has had a long and successful academic career as a renowned statistician and psykometrician, both at Uppsala University and at Princeton University and BI.