

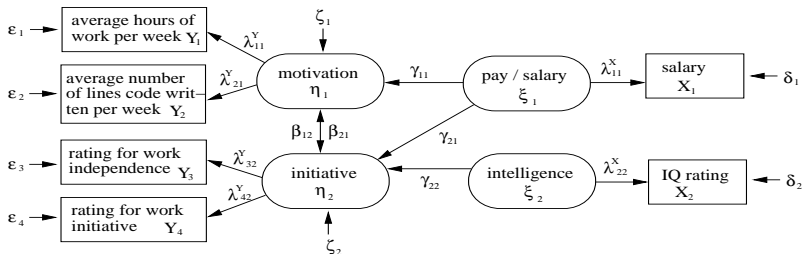
Structural Equation Modeling

Dr. Kerstin Hesse

Email: kerstin.hesse@hhl.de; Phone: +49 (0)341 9851-820; Office: HHL Main Building, Room 115A

HHL Leipzig Graduate School of Management, Jahnallee 59, 04109 Leipzig, Germany

Doctoral Program at HHL, June 1-2, 2012



Format of the Course

We will alternate between **introducing the new methods** and **practicing them on concrete examples**.

Software: Some Available Programs

- SPSS (licensed): factor analysis
- AMOS (licensed): LISREL
- SmartPLS (free, but registration required): PLS path modeling
- LISREL (licensed): LISREL
- EQS – Structural Equation Modeling Software (licensed): LISREL
- SAS PROC CALIS (licensed): LISREL
- SAS PROC PLS (licensed): PLS path modeling
- TETRAD (free): LISREL

Assessment: Take-Home Assignment, Handed Out After the Course

- **Submission deadline: Monday, July 09, 2012, 4:00 p.m.**
- Submission by email to me or in hard-copy handed in/sent to me.
- Rules of submission: You may collaborate with your colleagues (group work allowed), but you must prepare your own individual report.
- Format of submission: a typeset report or a neatly handwritten one.
- For email submission, please email [one pdf-file](#).

Help/Support: How to Get Help on the Take-Home Assignment

Contact me by email, phone or in person.

- Email: kerstin.hesse@hhl.de
- Phone: +49 (0)341 9851-820
- Office: HHL Main Building, Room 115A (I am usually there from 9:00 a.m. to 5:00 p.m., but please make an appointment by email.)

References on Factor Analysis

- K. Backhaus, B. Erichson, W. Plinke, R. Weiber: Multivariate Analysemethoden (13th edn.). Springer-Verlag, Berlin, Heidelberg, 2011.
- L. Fahrmeir, A. Hamerle, G. Tutz (eds.): Multivariate statistische Verfahren (2nd edn.). Walter de Gruyter, Berlin, 1996.
- A. Handl: Multivariate Analysemethoden: Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-PLUS (2nd edn.). Springer-Verlag, Heidelberg et al, 2010.
- W.K. Härdle, L. Simar: Applied Multivariate Statistical Analysis (3rd edn.). Springer-Verlag, Heidelberg et al, 2012.
- S. Sharma: Applied Multivariate Techniques. John Wiley & Sons, 1996.
- H. Rinne: Taschenbuch der Statistik (4th edn.). Verlag Harri Deutsch, Frankfurt am Main, 2008.
- B.G. Tabachnick, L.S. Fidell: Using Multivariate Statistics (5th edn.). Pearson Education, 2007.

- K. Backhaus, B. Erichson, R. Weiber: Fortgeschrittene Multivariate Analysemethoden (1st edn.). Springer-Verlag, Heidelberg et al, 2011.
- K.A. Bollen: Structural Equations with Latent Variables. New York, Wiley, 1989.
- L. Fahrmeir, A. Hamerle, G. Tutz (eds.): Multivariate statistische Verfahren (2nd edn.). Walter de Gruyter, Berlin, 1996.
- B.G. Tabachnick, L.S. Fidell: Using Multivariate Statistics (5th edn.). Pearson Education, 2007.
- R. Weiber, D. Mühlhaus: Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS and SPSS. Springer-Verlag, Berlin, Heidelberg, 2010.

- F. Bliemel, A. Eggert, G. Fassott, J. Henseler (eds.): Handbuch der PLS-Pfadmodellierung. Schäffer-Poeschel Verlag Stuttgart, 2005.
- T.K. Dijkstra: Latent variables and indices: Herman Wold's basic design and partial least squares. In: V.E. Vinzi et al. (eds.): Handbook of Partial Least Squares. Springer Handbooks of Computational Statistics 3, Springer-Verlag Berlin Heidelberg, 2010, Part 1, 23–46.
- M. Tenenhaus, V.E. Vinzia, Y.-M. Chatelin, C. Lauro: PLS path modeling. Computational Statistics & Data Analysis 48 (2005) 159-205.
- V.E. Vinzi et al. (eds.): Handbook of Partial Least Squares. Springer Handbooks of Computational Statistics 3, Springer-Verlag Berlin Heidelberg, 2010.

- V.E. Vinzi, L. Trinchera, S. Amato: PLS path modeling: From foundations to recent developments and open issues for model assessment and improvement. In: V.E. Vinzi et al. (eds.): Handbook of Partial Least Squares. Springer Handbooks of Computational Statistics 3, Springer-Verlag Berlin Heidelberg, 2010, Part 1, 47–82.
- R. Weiber, D. Mühlhaus: Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS and SPSS. Springer-Verlag, Berlin, Heidelberg, 2010.

Outline & Table of Contents

- 1 Revision of Linear Algebra and Variance/Covariance
 - Linear Algebra
 - Variance and Covariance
- 2 Factor Analysis
 - Idea of Factor Analysis
 - Main Theorem of Factor Analysis
 - Computing the Solution to the Problem of Factor Analysis
- 3 Introduction to Structural Equation Modeling (SEM)
 - Structural (Inner) Model
 - (Outer) Measurement Models
 - Approaches to Structural Equation Modeling (SEM)
- 4 LISREL (Linear Structural Relationships)
 - Setting Up the Structural Equation Model for LISREL
 - LISREL's Covariance Approach
 - Appendix: LISREL in Matrix Notation
- 5 PLS Path Modeling (Partial Least Squares Path Modeling)
 - Introduction to the PLS Path Model
 - Numerical Computation of the PLS Path Model: The PLS Algorithm

Structural Equation Modeling

Topic 1: Revision of Linear Algebra and Variance/Covariance

Dr. Kerstin Hesse

Email: kerstin.hesse@hhl.de; *Phone:* +49 (0)341 9851-820; *Office:* HHL Main Building, Room 115A

HHL Leipzig Graduate School of Management, Jahnallee 59, 04109 Leipzig, Germany

Doctoral Program at HHL, June 1-2, 2012

Topic 1: Revision of Linear Algebra & Variance/Covariance

● Linear Algebra

- general notation
- scalar multiplication of matrices and matrix multiplication
- determinant of a matrix
- eigenvalues and eigenvectors of a matrix
- transformation of a symmetric matrix to diagonal form

● Variance and Covariance

- random variables and samples
- expectation value and mean
- variance and estimating the variance from a sample
- centered and standardized random variables and data
- covariance and correlation coefficient, and estimating the covariance and the correlation coefficient from a sample
- formal manipulations of expectation values

General Notation: Scalars and Vectors

- **Scalar values (real numbers)** are denoted by lowercase letters: x, y, a, b, \dots
- **Random variables** are denoted by uppercase letters X, Y, Z, \dots . In **structural equation modeling** we will also use **Greek lowercase letters**, e.g. ξ or η , for random variables.
- **Vectors** (of real numbers or random variables) are denoted by lowercase boldface letters and are by default column vectors $\mathbf{x}, \mathbf{y}, \mathbf{w}, \dots$. In \mathbf{x}' the $'$ denotes taking the **transpose**.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} = (x_1, x_2, \dots, x_N)', \quad \mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{pmatrix} = (Y_1, Y_2, \dots, Y_p)'.$$

- The **length of a vector** $\mathbf{x} = (x_1, x_2, \dots, x_N)'$ is denoted $\|\mathbf{x}\|_2$ and is

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_N^2}.$$

General Notation: Matrices

- **Matrices** are denoted by boldface uppercase letters **A**, **B**, **X**, ...
- An $n \times m$ matrix **A** has n rows and m columns:

$$\mathbf{A} = (a_{i,j})_{\substack{i=1,2,\dots,n \\ j=1,2,\dots,m}} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix} \quad (1)$$

- The **entries of a matrix A** are usually denoted by the corresponding lowercase letter, i.e. $a_{i,j}$, with the **first index for the row** and the **second index for the column** (e.g. see (1)).
- Occasionally we may also use $\mathbf{A}_{i,j}$ to refer to the entry in the i th row and j th column of a matrix **A**.
- We may drop the comma between the indices in $a_{i,j}$ (i.e. write a_{ij} instead) if there is no ambiguity.

General Notation: Transpose of a Matrix

- For an $n \times m$ matrix \mathbf{A} (given by (1)), \mathbf{A}' denotes the transpose of the matrix \mathbf{A} which is the $m \times n$ matrix given by

$$\mathbf{A}' = \begin{pmatrix} a_{1,1} & a_{2,1} & \dots & a_{n,1} \\ a_{1,2} & a_{2,2} & \dots & a_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,m} & a_{2,m} & \dots & a_{n,m} \end{pmatrix}. \quad (2)$$

The entries of the i th row of \mathbf{A} become the entries of the i th column of \mathbf{A}' (indicated in (1) and (2) in violet for the 1st row/column).

- Example of a 2×3 matrix and its transpose:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad \mathbf{A}' = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

Here the entries of \mathbf{A} are $a_{1,1} = 1$, $a_{1,2} = 2$, $a_{1,3} = 3$, $a_{2,1} = 4$, $a_{2,2} = 5$ and $a_{2,3} = 6$.

Multiplication of Vectors and Matrices by Scalars

Scalar Multiplication: Matrices and vectors are multiplied by a scalar (real number) by multiplying each entry of the matrix and vector respectively by the scalar:

$$\alpha \cdot \mathbf{x} = \begin{pmatrix} \alpha \cdot x_1 \\ \alpha \cdot x_2 \\ \vdots \\ \alpha \cdot x_n \end{pmatrix}, \quad \beta \cdot \mathbf{A} = \begin{pmatrix} \beta \cdot a_{1,1} & \beta \cdot a_{1,2} & \dots & \beta \cdot a_{1,m} \\ \beta \cdot a_{2,1} & \beta \cdot a_{2,2} & \dots & \beta \cdot a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta \cdot a_{n,1} & \beta \cdot a_{n,2} & \dots & \beta \cdot a_{n,m} \end{pmatrix}.$$

Example of Scalar Multiplication:

$$3 \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix}, \quad (-2) \cdot \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} -2 & -4 & -6 \\ -8 & -10 & -12 \\ -14 & -16 & -18 \end{pmatrix}.$$

Matrix Multiplication

An $n \times m$ matrix $\mathbf{A} = (a_{ik})_{\substack{i=1,\dots,n \\ k=1,\dots,m}}$ can be multiplied with an $m \times p$ matrix $\mathbf{B} = (b_{kj})_{\substack{k=1,\dots,m \\ j=1,\dots,p}}$ to form the $n \times p$ matrix $\mathbf{A} \cdot \mathbf{B} = \mathbf{A} \mathbf{B} = \mathbf{C} = (c_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,p}}$ in the following way:

$$\begin{aligned} c_{ij} &= \sum_{k=1}^m a_{ik} \cdot b_{kj} = a_{i1} \cdot b_{1j} + a_{i2} \cdot b_{2j} + \dots + a_{im} \cdot b_{mj} \\ &= \left(\begin{array}{l} \text{inner/scalar product of the } i\text{th row vector of } \mathbf{A} \\ \text{with the } j\text{th column vector of } \mathbf{B} \end{array} \right) \end{aligned}$$

Beware: Matrix multiplication is **not commutative**, i.e. usually $\mathbf{A} \mathbf{B} \neq \mathbf{B} \mathbf{A}$ (if both $\mathbf{A} \mathbf{B}$ and $\mathbf{B} \mathbf{A}$ are defined).

Ex. 1.1 (Matrix Multiplication): Execute the **matrix multiplication**

$$\begin{pmatrix} 1 & 1 & -1 \\ -1 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Determinants of 2×2 and 3×3 Matrices

The **determinant** $\det(\mathbf{A})$ of the 2×2 matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is given by

$$\det(\mathbf{A}) = a \cdot d - c \cdot b.$$

The **determinant** $\det(\mathbf{A})$ of the 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$$

is given by

$$\begin{aligned} \det(\mathbf{A}) = & a_{1,1} \cdot a_{2,2} \cdot a_{3,3} + a_{1,2} \cdot a_{2,3} \cdot a_{3,1} + a_{1,3} \cdot a_{2,1} \cdot a_{3,2} \\ & - a_{3,1} \cdot a_{2,2} \cdot a_{1,3} - a_{3,2} \cdot a_{2,3} \cdot a_{1,1} - a_{3,3} \cdot a_{2,1} \cdot a_{1,2}. \end{aligned}$$

Determinants of $n \times n$ Matrices with $n \geq 3$

We compute the determinant of an $n \times n$ matrix $\mathbf{A} = (a_{ij})_{i,j=1,\dots,n}$

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i,1} & a_{i,2} & \dots & a_{i,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix}$$

by expanding with respect to the 1st column:

$$\det(\mathbf{A}) = \sum_{j=1}^n (-1)^{j+1} \cdot a_{j1} \cdot \det(\mathbf{C}_{j1}),$$

where \mathbf{C}_{j1} is the $(n-1) \times (n-1)$ matrix obtained by deleting from \mathbf{A} the 1st column and the i th row.

Exercises: Determinants of Matrices

Ex. 1.2 (Determinants of 2×2 and 3×3 Matrices):

Compute the **determinants** of the following two matrices

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Ex. 1.3 (General Formula for the Determinant):

Use the **expansion with respect to the first column formula** from page 17 to compute the **determinant** of

$$\mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Eigenvalues and Eigenvectors

Let \mathbf{A} be an $n \times n$ matrix. A real number λ is called an **eigenvalue of \mathbf{A}** if there exists a **non-zero vector \mathbf{x}** such that

$$\mathbf{A} \mathbf{x} = \lambda \mathbf{x}. \quad (3)$$

\mathbf{x} is then called an **eigenvector of \mathbf{A}** corresponding to the eigenvalue λ .

Interpretation: (3) means that \mathbf{x} is mapped onto a **multiple** of itself.

To find the n eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ and the corresponding **eigenvectors** of an $n \times n$ matrix \mathbf{A} :

- 1 Compute the **characteristic polynomial** (of degree n in λ)

$$p(\mathbf{A}, \lambda) = \det(\lambda \mathbf{I} - \mathbf{A}) = 0 \quad (\mathbf{I} \text{ is the identity matrix.})$$

and **find the zeros/roots** to get the n eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of \mathbf{A} .

- 2 For each eigenvalue λ_j **solve the linear system**

$$\mathbf{A} \mathbf{x}_j = \lambda_j \mathbf{x}_j \quad \text{or equivalently} \quad (\lambda_j \mathbf{I} - \mathbf{A}) \mathbf{x}_j = \mathbf{0}$$

to find the eigenvectors \mathbf{x}_j corresponding to λ_j .

Ex. 1.4: Eigenvectors and Eigenvalues

Consider the **symmetric matrix**

$$\mathbf{A} = \begin{pmatrix} \frac{3}{2} & 0 & \frac{1}{2} \\ 0 & 3 & 0 \\ \frac{1}{2} & 0 & \frac{3}{2} \end{pmatrix},$$

- 1 Compute the **eigenvalues** $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and the **corresponding eigenvectors** $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ of \mathbf{A} (where $\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i$, $i = 1, 2, 3$).
- 2 **Wait with this part for the next two slides until we have discussed this:** Find an **orthogonal matrix** \mathbf{S} such that

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{S}'\mathbf{A}\mathbf{S} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}, \quad (4)$$

with $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Execute the matrix multiplication in (4) to verify that you have chosen \mathbf{S} correctly.

Transformation of a Symmetric Matrix to Diagonal Form I

Let \mathbf{A} be a real **symmetric** $n \times n$ matrix, i.e. $\mathbf{A} = \mathbf{A}'$.

- (i) Then \mathbf{A} has n **real** eigenvalues (counted with multiplicities).
 - (ii) The eigenvectors to different eigenvalues are **orthogonal**.
 - (iii) For an eigenvalue with **multiplicity** $k > 1$ (i.e. it appears k -times as a root of the characteristic polynomial) we find k **linearly independent eigenvectors**. These k eigenvectors can be chosen **orthogonal (perpendicular)** to each other.
-

Consequence: A real **symmetric matrix** has n **real eigenvalues** $\lambda_1, \lambda_2, \dots, \lambda_n$ and n corresponding **orthogonal eigenvectors** $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$.

In addition, we may demand that $\|\mathbf{x}_1\|_2 = \|\mathbf{x}_2\|_2 = \dots = \|\mathbf{x}_n\|_2 = 1$, i.e. the eigenvectors are **normalized** (have all **length one**).

Transformation of a Symmetric Matrix to Diagonal Form II

Let \mathbf{A} be a real symmetric $n \times n$ matrix, i.e. $\mathbf{A} = \mathbf{A}'$, and let $\lambda_1, \lambda_2, \dots, \lambda_n$ be its n real eigenvalues, and let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be n corresponding normalized orthogonal eigenvectors.

From the equations $\mathbf{A} \mathbf{x}_1 = \lambda_1 \mathbf{x}_1, \dots, \mathbf{A} \mathbf{x}_n = \lambda_n \mathbf{x}_n$, we have

$$\mathbf{A} \underbrace{\begin{pmatrix} \vdots & \vdots & \vdots \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ \vdots & \vdots & \vdots \end{pmatrix}}_{=\mathbf{S}} = \underbrace{\begin{pmatrix} \vdots & \vdots & \vdots \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ \vdots & \vdots & \vdots \end{pmatrix}}_{=\mathbf{S}} \underbrace{\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}}_{=\mathbf{D}},$$

and the matrix \mathbf{S} is orthogonal, i.e. $\mathbf{S}^{-1} = \mathbf{S}'$.

Multiplying $\mathbf{A} \mathbf{S} = \mathbf{S} \mathbf{D}$ from the right with $\mathbf{S}^{-1} = \mathbf{S}'$, we find

$$\mathbf{S}' \mathbf{A} \mathbf{S} = \underbrace{\mathbf{S}' \mathbf{S}}_{=\mathbf{I}} \mathbf{D} = \mathbf{I} \mathbf{D} = \mathbf{D} \quad \text{and thus} \quad \mathbf{S}' \mathbf{A} \mathbf{S} = \mathbf{D}.$$

Random Variables and Samples

Let X and Y be **random variables**. (A random variable is a feature that can be sampled at objects and whose value depends on chance.)

In this course, all random variables are **metric**, i.e. the values of the random variable are **real numbers** and the **difference of two values of the random variable has meaning**.

Examples of Metric Random Variables: price of a product, height, temperature, mark in percentage.

Random variables are measured in **populations**.

Sample: A **sample** is a randomly drawn collection of objects e_1, e_2, \dots, e_n from the population on which we measure X : $x_j = \text{value of } X \text{ on object } e_j$.

Example: $X =$ mark of students in percentage. Population = all students. Sample of 3 students e_1, e_2, e_3 with marks $x_1 = 80, x_2 = 90, x_3 = 70$.

Expectation Value and Mean

Expectation Value: The expectation value $\mu_X = E(X)$ of a random variable X is the value that X assumes on average.

Mean (Value): In a sample of objects e_1, e_2, \dots, e_n with values x_1, x_2, \dots, x_n for the random variable X , the expectation value $E(X)$ is estimated by the mean (value):

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Example: X = mark of students in percentage. Sample of 3 students with marks $x_1 = 80$, $x_2 = 90$ and $x_3 = 70$. $E(X)$ is estimated by

$$\bar{x} = \frac{1}{3} (x_1 + x_2 + x_3) = \frac{1}{3} (80 + 90 + 70) = 80.$$

Variance and Standard Deviation of a Random Variable

Variance: The variance $\text{Var}(X) = \sigma_X^2$ of a random variable X is defined by

$$\text{Var}(X) = E([X - E(X)]^2) = E(X^2) - [E(X)]^2$$

and describes the (squared) average variation of the values of X around the expectation value $E(X)$.

Standard Deviation: The standard deviation

$$\sigma_X = \sqrt{\text{Var}(X)}$$

describes the average variation of the values of X around $E(X)$.

Estimating Variance and Standard Deviation

Estimating $\text{Var}(X)$ and σ_X : In a sample of objects e_1, e_2, \dots, e_n with values x_1, x_2, \dots, x_n for the random variable X , the **variance** $\text{Var}(X) = \sigma_X^2$ and the **standard deviation** σ_X are estimated by the **empirical variance** $\widehat{\sigma}_X^2 = s_X^2$ the **empirical standard deviation** $\widehat{\sigma}_X = s_X$, defined by

$$\widehat{\sigma}_X^2 = s_X^2 = \frac{1}{n-1} [(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2] = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

and $\widehat{\sigma}_X = s_X = \sqrt{s_X^2}$, respectively.

Example: X = mark of students in percentage. Sample of 3 students with marks $x_1 = 80$, $x_2 = 90$ and $x_3 = 70$. $E(X)$ is estimated by $\bar{x} = 80$. Hence,

$$\widehat{\sigma}_X^2 = s_X^2 = \frac{1}{3-1} [(80 - 80)^2 + (90 - 80)^2 + (70 - 80)^2] = \frac{200}{2} = 100$$

and $\widehat{\sigma}_X = s_X = \sqrt{100} = 10$.

Centered and Standardized Random Variables

By defining for a random variable X with $E(X) = \mu_X$ and $\text{Var}(X) = \sigma_X^2$

$$W = X - E(X) = X - \mu_X \quad \text{and} \quad Z = \frac{X - E(X)}{\sigma_X} = \frac{X - \mu_X}{\sigma_X} \quad (5)$$

we obtain:

- a centered random variable W with $E(W) = 0$ and $\text{Var}(W) = \sigma_X^2$,
- a standardized random variable Z with $E(Z) = 0$ and $\text{Var}(Z) = 1$

Centering and Standardizing Data from a Sample: In a sample of objects e_1, e_2, \dots, e_n with values x_1, x_2, \dots, x_n for the random variable X , with mean value \bar{x} and empirical standard deviation s_X we compute centered data and standardized data via

$$w_i = x_i - \bar{x} \quad \text{and} \quad z_i = \frac{x_i - \bar{x}}{s_X}, \quad \text{respectively.}$$

Example: Centered and Standardized Data

Example: X = mark of students in percentage. Sample of 3 students with marks $x_1 = 80$, $x_2 = 90$ and $x_3 = 70$. We found $\bar{x} = 80$ and $s_X = 10$.

The **centered data** of the centered random variable $W = X - \mu_X$ is:

$$w_1 = x_1 - \bar{x} = 80 - 80 = 0$$

$$w_2 = x_2 - \bar{x} = 90 - 80 = 10$$

$$w_3 = x_3 - \bar{x} = 70 - 80 = -10$$

The **standardized data** of the standardized random variable $Z = (X - \mu_X)/\sigma_X$ is given by:

$$z_1 = \frac{x_1 - \bar{x}}{s_X} = \frac{80 - 80}{10} = 0$$

$$z_2 = \frac{x_2 - \bar{x}}{s_X} = \frac{90 - 80}{10} = 1$$

$$z_3 = \frac{x_3 - \bar{x}}{s_X} = \frac{70 - 80}{10} = -1$$

Covariance of Random Variables

Covariance: The **covariance** $\text{Cov}(X, Y)$ of two random variables X and Y is defined by

$$\text{Cov}(X, Y) = E([X - E(X)] \cdot [Y - E(Y)]).$$

The covariance **measures the correlation of the two random variables.**

Estimating the Covariance from a Sample: In a sample of objects e_1, e_2, \dots, e_n with values x_1, x_2, \dots, x_n and the mean \bar{x} for the random variable X and values y_1, y_2, \dots, y_n and the mean \bar{y} for the random variable Y , the **covariance** $\text{Cov}(X, Y)$ is estimated by the **empirical covariance** $\widehat{\text{Cov}}(X, Y)$, defined by

$$\begin{aligned}\widehat{\text{Cov}}(X, Y) &= \frac{1}{n-1} [(x_1 - \bar{x}) \cdot (y_1 - \bar{y}) + \dots + (x_n - \bar{x}) \cdot (y_n - \bar{y})] \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})\end{aligned}$$

Correlation Coefficient: A scale-independent measure of the correlation of two random variables X and Y is given by the correlation coefficient

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Estimating the Correlation Coefficient from a Sample: The correlation coefficient is estimated by the empirical correlation coefficient

$$\hat{\rho}(X, Y) = \frac{\widehat{\text{Cov}}(X, Y)}{s_X \cdot s_Y}$$

where $\widehat{\text{Cov}}(X, Y)$, s_X and s_Y are the empirical covariance and empirical variances of X and Y , respectively, computed from the sample.

Ex. 1.5: Mean, Variance, Covariance and Correlation

Consider the random variables $X = \text{mark of students in percentage}$ and $Y = \text{age of the student}$. In a sample of 3 students we found the values

$$x_1 = 80, x_2 = 90, x_3 = 70 \quad \text{and} \quad y_1 = 24, y_2 = 23, y_3 = 22$$

for X and Y , respectively. Estimate the covariance and the correlation coefficient of X and Y from the sample.

Manipulating Expectation Value, Variance & Covariance

The **expectation value** is **linear**, i.e. for two **random variables X and Y** and **real numbers a and b** we have

$$E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y). \quad (6)$$

Ex. 1.6 (Formal Manipulations of Expectation Values):

Let X , Y and W be random variables with expectation values $\mu_X = E(X)$, $\mu_Y = E(Y)$ and $\mu_W = E(W)$ and standard deviations σ_X , σ_Y and σ_W , respectively. Let a , b and c be real numbers. Use (6) to **verify the following relations**:

$$\begin{aligned} E\left([a \cdot (X - \mu_X) + b \cdot (Y - \mu_Y)] \cdot [c \cdot (W - \mu_W)]\right) \\ = a \cdot c \cdot \text{Cov}(X, W) + b \cdot c \cdot \text{Cov}(Y, W), \end{aligned}$$

$$\text{Var}(a \cdot X) = a^2 \cdot \text{Var}(X).$$

Structural Equation Modeling

Topic 2: Factor Analysis

Dr. Kerstin Hesse

Email: kerstin.hesse@hhl.de; *Phone:* +49 (0)341 9851-820; *Office:* HHL Main Building, Room 115A

HHL Leipzig Graduate School of Management, Jahnallee 59, 04109 Leipzig, Germany

Doctoral Program at HHL, June 1-2, 2012

Topic 2: Factor Analysis

- Idea of Factor Analysis
 - idea of factor analysis
 - introductory example
- Main Theorem of Factor Analysis
 - model of factor analysis
 - correlation matrix
 - main theorem of factor analysis
 - communalities and reduced correlation matrix
- Computing the Solution to the Problem of Factor Analysis
 - computation of the factor loading matrix
 - computing the values of the factors
 - interpretation of the values of the factors
- Outlook: Factor Rotation

Idea of Factor Analysis

Aim: The aim of **factor analysis** is to **explain m measured variables through a few common factors** in a sufficiently accurate but simple way.

Factor analysis seeks to represent the information contained in the measured variables in a **more concise way through fewer common factors**.

Example (Intelligence Tests):

- A large number of questions have to be answered by the test persons.
 - Can the responses be explained through the **one factor intelligence**?
Or should they be explained through **a number of different abilities (factors)**, e.g. linguistic aptitude, mathematical aptitude, memory, abstract thinking, etc.?
-

Note: There has to be some **compromise** in factor analysis, as **more simplification means usually less accuracy**.

Example & Problem for our Toy Exercise

Problem: Stiftung Warentest evaluates 5 types of cereal: e_1, e_2, \dots, e_5 .

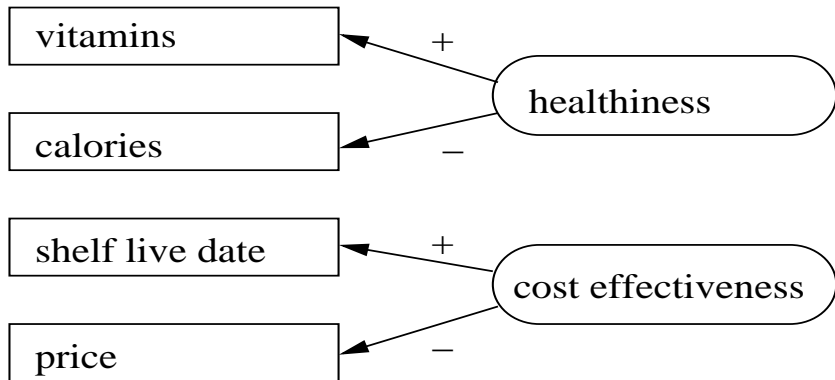
For each type of cereal, Stiftung Warentest does a chemical analysis and determines the average market price and the average shelf live.

For comparability the results are converted into a metric rating from 1 to 5 for the following categories:

- **Vitamins:** The higher the rating the higher the vitamin content.
- **Calories:** The higher the rating the higher the number of calories.
- **Shelf live/Sell-by date:** The higher the rating the longer the shelf live.
- **Price:** The higher the rating the higher the price.

Research Question: Can these ratings be explained by some common factors F_1, F_2, \dots , such as healthiness and cost effectiveness?

Possible Diagram for Example & Toy Exercise



The **plus and minus signs** at the arrows have to be read as follows:

The higher the healthiness the higher the (rating for) vitamin count.

The higher the healthiness the lower the (rating the) calorie count.

The higher the cost effectiveness the higher the (rating for) shelf live.

The higher the cost effectiveness the lower the (rating for) price.

Possible Model Equations for Example & Toy Exercise

Metric Measured Variables:

- X_1 = rating for vitamins
- X_2 = rating for calories
- X_3 = rating for shelf live date
- X_4 = rating for price

Factors/Metric Latent Variables:

- F_1 = healthiness,
- F_2 = cost effectiveness

Errors/Effects not Reflected in the Model: U_1, U_2, U_3, U_4

Model Equations for Diagram:

We formulate our model equations for the **standardized variables**

$$Z_j = (X_j - E(X_j))/\sigma_{X_j}.$$

$$Z_1 = a_{1,1} \cdot F_1 + U_1$$

$$Z_2 = a_{2,1} \cdot F_1 + U_2$$

$$Z_3 = a_{3,2} \cdot F_2 + U_3$$

$$Z_4 = a_{4,2} \cdot F_2 + U_4$$

Based on the diagram we expect $a_{1,1} > 0$, $a_{2,1} < 0$, $a_{3,2} > 0$, $a_{4,2} < 0$.

Generally, for each standardized metric variable we postulate the model:

$$Z_j = a_{j1} \cdot F_1 + a_{j2} \cdot F_2 + \dots + a_{jp} \cdot F_p + U_j, \quad j = 1, 2, 3, 4$$

Model Equations for Example & Toy Exercise

Model: For the standardized variables $Z_j = (X_j - E(X_j))/\sigma_{X_j}$, we postulate a linear model with $p < 4$ factors F_1, F_2, \dots, F_p

$$Z_j = a_{j1} \cdot F_1 + a_{j2} \cdot F_2 + \dots + a_{jp} \cdot F_p + U_j, \quad j = 1, 2, \dots, 4, \quad (7)$$

that explains Z_j in terms of the unknown common factors F_1, F_2, \dots, F_p .

- $a_{j1}, a_{j2}, \dots, a_{jp}$ are the coefficients (called factor loadings) in (7).
- U_1, U_2, \dots, U_p are called unique factors and cover random errors and effects not explained by the common factors.

Important:

- The number p of factors and the factors themselves are unknown, and it is the aim of exploratory factor analysis to determine these!
- After the factor analysis, we have to interpret the factors!
- In the diagram and model equations on pages 37 and 38 we have shown one possible result of the factor analysis for our toy example.

Matrix Model Equations for Example & Toy Exercise

We write (7) as a **matrix equation** $\mathbf{z} = \mathbf{A}\mathbf{f} + \mathbf{u}$ ($\Leftrightarrow \mathbf{z}' = \mathbf{f}'\mathbf{A}' + \mathbf{u}'$) with

$$\mathbf{z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_4 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,p} \\ \vdots & & \vdots \\ a_{4,1} & \cdots & a_{4,p} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} F_1 \\ \vdots \\ F_p \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} U_1 \\ \vdots \\ U_4 \end{pmatrix}.$$

For the i th type of cereal e_i , we have a **standardized data vector**

$$\mathbf{z}'_i = (z_{i1}, z_{i2}, \dots, z_{i4})$$

and a corresponding **common factor value vector** \mathbf{f}'_i and **unique factor value vector** \mathbf{u}'_i

$$\mathbf{f}'_i = (f_{i1}, f_{i2}, \dots, f_{ip}) \quad \text{and} \quad \mathbf{u}'_i = (u_{i1}, u_{i2}, \dots, u_{i4})$$

and hence a **linear system (of equations)**

$$\mathbf{z}_i = \mathbf{A}\mathbf{f}_i + \mathbf{u}_i \quad \Leftrightarrow \quad \mathbf{z}'_i = \mathbf{f}'_i\mathbf{A}' + \mathbf{u}'_i. \quad (8)$$

Linear System for Example & Toy Exercise

From (8) we get the **linear system (of equations)** for the data:

$$\mathbf{Z} = \mathbf{F} \mathbf{A}' + \mathbf{U}$$

where \mathbf{A} is the **factor loading matrix** defined on page 40 and

$$\mathbf{Z} = \begin{pmatrix} z_{1,1} & \cdots & z_{1,4} \\ \vdots & & \vdots \\ z_{5,1} & \cdots & z_{5,4} \end{pmatrix} \leftarrow \begin{array}{l} \mathbf{z}'_1 = \text{standardized data for cereal } e_1 \\ \vdots \\ \mathbf{z}'_5 = \text{standardized data for cereal } e_5 \end{array}$$

$$\mathbf{F} = \begin{pmatrix} f_{1,1} & \cdots & f_{1,p} \\ \vdots & & \vdots \\ f_{5,1} & \cdots & f_{5,p} \end{pmatrix} \leftarrow \begin{array}{l} \mathbf{f}'_1 = \text{values of common factors for cereal } e_1 \\ \vdots \\ \mathbf{f}'_5 = \text{values of common factors for cereal } e_5 \end{array}$$

$$\mathbf{U} = \begin{pmatrix} u_{1,1} & \cdots & u_{1,4} \\ \vdots & & \vdots \\ u_{5,1} & \cdots & u_{5,4} \end{pmatrix} \leftarrow \begin{array}{l} \mathbf{u}'_1 = \text{values of unique factors for cereal } e_1 \\ \vdots \\ \mathbf{u}'_5 = \text{values of unique factors for cereal } e_5 \end{array}$$

Ex. 2.1: Standardized Data in Example & Toy Exercise

Given the rating for vitamins ($= X_1$), rating for calories ($= X_2$), rating for shelf live date ($= X_3$) and rating for price ($= X_4$) for 5 types of cereal in the following table, compute the data for the corresponding **standardized variables** Z_1, \dots, Z_4 and write down the **standardized data matrix**:

Cereal	X_1 (Vitamins)	X_2 (Calories)	X_3 (Shelf Live)	X_4 (Price)
e_1	4	2	3	3
e_2	2	4	3	3
e_3	3	3	3	3
e_4	3	3	2	4
e_5	3	3	4	2

Ex. 2.2: Some Model Equations in the Toy Exercise

Write down the **model equations** for each random variable X_j for cereals e_1 and e_2 . Inspect the model equations:

- What are the **unknowns**?
- Compare the model equations with the equations in **(multiple regression)**. Where does the **difference** lie?

Observation for the Example & Toy Exercise

Our linear system in the toy example

$$\underbrace{\mathbf{Z}}_{5 \times 4} = \underbrace{\mathbf{F}}_{5 \times p} \underbrace{\mathbf{A}'}_{p \times 4} + \underbrace{\mathbf{U}}_{5 \times 4}$$

does usually **not** have square matrices; hence the given matrices **cannot be invertible**.

Problem: We only know \mathbf{Z} , but need to determine everything else!

Approach: We need to **invest more information about the given data**:

- If the standardized variables Z_1, Z_2, \dots, Z_4 can be explained by a few common factors, then Z_1, Z_2, \dots, Z_4 **must be correlated**.
- We will use the **correlation matrix** of our data for Z_1, Z_2, \dots, Z_4 to compute the number of factors p , the factor loading matrix \mathbf{A} and the factor value matrix \mathbf{F} .

Factor Analysis: Setup & Model for the Random Variables

Given: Metric variables X_1, X_2, \dots, X_m that are correlated. Z_i is the standardized variable corresponding to X_i , $i = 1, 2, \dots, m$.

Aim: X_1, X_2, \dots, X_m shall be explained through a few common factors F_1, F_2, \dots, F_p with $p < m$.

Model: The postulated linear model is

$$Z_j = a_{j1} \cdot F_1 + a_{j2} \cdot F_2 + \dots + a_{jp} \cdot F_p + U_j = \sum_{k=1}^p a_{jk} \cdot F_k + U_j, \quad (9)$$

$j = 1, 2, \dots, m$, where:

- The U_j is the unique factor of Z_j that captures random variation and effects not represented by the common factors.
- The coefficients a_{j1}, \dots, a_{jp} are the factor loadings. They are unique for each random variable Z_j .

Factor Analysis: Setup & Model for the Sampled Data I

Sampled Data: data for X_1, X_2, \dots, X_m for n objects e_1, e_2, \dots, e_n :
 $x_{j1}, x_{j2}, \dots, x_{jm}$ = data for X_1, X_2, \dots, X_m for object e_j

Standardized Data: $z_{j1}, z_{j2}, \dots, z_{jm}$ is the corresponding standardized data for object e_j ,

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad \bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{kj}, \quad s_j = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}$$

for the standardized variables $Z_j = (X_j - E(X_j))/\sigma_{X_j}$, $j = 1, 2, \dots, m$.

Standardized Data Matrix:

$$\mathbf{Z} = \begin{pmatrix} z_{1,1} & \cdots & z_{1,m} \\ \vdots & & \vdots \\ z_{n,1} & \cdots & z_{n,m} \end{pmatrix} \begin{array}{l} \leftarrow \text{standardized data for object } e_1 \\ \vdots \\ \leftarrow \text{standardized data for object } e_n \end{array}$$
$$\begin{array}{ccc} \uparrow & & \uparrow \\ Z_1 & \cdots & Z_m \end{array}$$

Factor Analysis: Setup & Model for the Sampled Data II

Model: The linear model (9) results in a **matrix equation** for the data:

$$\underbrace{\mathbf{Z}}_{n \times m} = \underbrace{\mathbf{F}}_{n \times p} \underbrace{\mathbf{A}'}_{p \times m} + \underbrace{\mathbf{U}}_{n \times m}$$

$$\mathbf{F} = \begin{pmatrix} f_{1,1} & \cdots & f_{1,p} \\ \vdots & & \vdots \\ f_{n,1} & \cdots & f_{n,p} \end{pmatrix} \begin{array}{l} \leftarrow \text{values of common factors for object } e_1 \\ \vdots \\ \leftarrow \text{values of common factors for object } e_n \end{array}$$

$\begin{array}{ccc} \uparrow & & \uparrow \\ F_1 & \cdots & F_p \end{array}$

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,p} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,p} \end{pmatrix} \begin{array}{l} \leftarrow \text{factor loadings for random variable } Z_1 \\ \vdots \\ \leftarrow \text{factor loadings for random variable } Z_m \end{array}$$

$\begin{array}{ccc} \uparrow & & \uparrow \\ F_1 & \cdots & F_p \end{array}$

$$\mathbf{U} = \begin{pmatrix} u_{1,1} & \cdots & u_{1,m} \\ \vdots & & \vdots \\ u_{n,1} & \cdots & u_{n,m} \end{pmatrix} \begin{array}{l} \leftarrow \text{values of unique factors for object } e_1 \\ \vdots \\ \leftarrow \text{values of unique factors for object } e_n \end{array}$$
$$\begin{array}{ccc} \uparrow & & \uparrow \\ Z_1 & \dots & Z_m \end{array}$$

Assumptions on the Factor Analysis Model:

- The factors F_j are standardized, i.e. $E(F_j) = 0$ and $\text{Var}(F_j) = 1$ for all $j = 1, 2, \dots, p$.
- The unique factor are centered, i.e. $E(U_j) = 0$ for all $j = 1, 2, \dots, m$.
- U_i and U_j are uncorrelated, i.e. $\text{Cov}(U_i, U_j) = 0$ if $i \neq j$.
- F_i and F_j are uncorrelated, i.e. $\text{Cov}(F_i, F_j) = 0$ if $i \neq j$.
- U_i and F_j are uncorrelated, i.e. $\text{Cov}(U_i, F_j) = 0$.

Correlation Matrix of the Data

For the **standardized data matrix**

$$\underbrace{\mathbf{Z}}_{n \times m} = \begin{pmatrix} z_{1,1} & \cdots & z_{1,m} \\ \vdots & & \vdots \\ z_{n,1} & \cdots & z_{n,m} \end{pmatrix} \quad \text{with} \quad \underbrace{\mathbf{Z}'}_{m \times n} = \begin{pmatrix} z_{1,1} & \cdots & z_{n,1} \\ \vdots & & \vdots \\ z_{1,m} & \cdots & z_{n,m} \end{pmatrix}$$

the **correlation matrix** $\mathbf{R} = (r_{ij})_{i,j=1,2,\dots,m}$ of the data of the random variables X_1, X_2, \dots, X_m is given by:

$$\mathbf{R} = \frac{1}{n-1} \mathbf{Z}' \mathbf{Z}$$

From matrix multiplication, the (i,j) th entry of $\mathbf{R} = (r_{ij})_{i,j=1,\dots,m}$ is

$$r_{ij} = \frac{1}{n-1} \sum_{k=1}^n z_{ki} \cdot z_{kj}$$

Understanding the Correlation Matrix of the Data

The entry r_{ij} of the correlation matrix \mathbf{R} is really an estimator for the correlation coefficient $\varrho(X_i, X_j)$ of X_i and X_j :

$$\begin{aligned} r_{ij} &= \frac{1}{n-1} \sum_{k=1}^n z_{ki} \cdot z_{kj} = \frac{1}{n-1} \sum_{k=1}^n \frac{x_{ki} - \bar{x}_i}{s_i} \cdot \frac{x_{kj} - \bar{x}_j}{s_j} \\ &= \frac{\frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i) \cdot (x_{kj} - \bar{x}_j)}{s_i \cdot s_j} \\ &= \begin{cases} \frac{s_i^2}{s_i \cdot s_i} = 1 = \text{Var}(Z_i) & \text{if } i = j, \\ \frac{\widehat{\text{Cov}}(X_i, X_j)}{s_i \cdot s_j} = \widehat{\varrho}(X_i, X_j) & \text{if } i \neq j. \end{cases} \end{aligned}$$

Ex. 2.3 (Correlation Matrix): Compute the correlation matrix for our toy example.

Test Whether Correlations are Significant

Important: A factor analysis makes only sense if the variables X_1, X_2, \dots, X_m are **correlated**, that is, if **some of the entries r_{ij} , $i \neq j$, of the correlation matrix \mathbf{R} differ significantly from zero.**

- r_{ij} models the **correlation coefficient** of X_i and X_j defined by

$$\rho_{ij} = \rho(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sigma_{X_i} \cdot \sigma_{X_j}}$$

- Hypothesis testing to determine whether $\rho_{ij} \neq 0$:

Null hypothesis H_0 : $\rho_{ij} = 0$

Alternative hypothesis H_1 : $\rho_{ij} \neq 0$

- $t = \frac{\sqrt{n-2} r_{ij}}{\sqrt{1-r_{ij}^2}}$ follows a t -distribution with $n-2$ degrees of freedom.
- For the significance level α , determine the value $t_{n-2, \alpha/2}$ from the t -distribution: **Reject H_0 , if $|t| > t_{n-2, \alpha/2}$; otherwise accept H_0 .**

Derivation of the Main Theorem of Factor Analysis I

$$\text{Model: } \mathbf{Z} = \mathbf{F}\mathbf{A}' + \mathbf{U} \quad \Leftrightarrow \quad \mathbf{Z}' = \mathbf{A}\mathbf{F}' + \mathbf{U}' \quad (10)$$

$$\text{Correlation Matrix: } \mathbf{R} = \frac{1}{n-1} \mathbf{Z}'\mathbf{Z} \quad (11)$$

In (11) substitute \mathbf{Z}' and \mathbf{Z} by the equations (10) of our model:

$$\begin{aligned} \mathbf{R} &= \frac{1}{n-1} (\mathbf{A}\mathbf{F}' + \mathbf{U}')(\mathbf{F}\mathbf{A}' + \mathbf{U}) \\ &= \frac{1}{n-1} \left(\mathbf{A} \underbrace{\mathbf{F}'\mathbf{F}}_{=(n-1)\mathbf{I}} \mathbf{A}' + \underbrace{\mathbf{A}\mathbf{F}'\mathbf{U}}_{=0} + \underbrace{\mathbf{U}'\mathbf{F}\mathbf{A}'}_{=0} + \underbrace{\mathbf{U}'\mathbf{U}}_{=(n-1)\boldsymbol{\Psi}} \right) \\ &= \mathbf{A}\mathbf{A}' + \boldsymbol{\Psi} \end{aligned}$$

- $\mathbf{F}'\mathbf{F} = (n-1)\mathbf{I}$ follows from $\text{Var}(F_j) = 1$ and $\text{Cov}(F_i, F_j) = 0$ if $i \neq j$. (\mathbf{I} is the identity matrix with 1 on the diagonal and 0 everywhere else.)
- $\mathbf{F}'\mathbf{U} = \mathbf{0}$ and $\mathbf{U}'\mathbf{F} = \mathbf{0}$ follow from $\text{Cov}(F_i, U_j) = 0$ for all i and j .
- $\boldsymbol{\Psi} = \frac{1}{n-1} \mathbf{U}'\mathbf{U}$ is the covariance matrix of the data for U_1, U_2, \dots, U_m .

Derivation of the Main Theorem of Factor Analysis II

$\frac{1}{n-1} \mathbf{F}' \mathbf{F} = \mathbf{I}$ is the covariance matrix of the factors F_1, F_2, \dots, F_p , and it equals the identity matrix as the factors are standardized and uncorrelated.

Because of $\text{Cov}(U_i, U_j) = 0$ if $i \neq j$, the covariance matrix Ψ of the unique factors U_1, U_2, \dots, U_m is diagonal with diagonal entries $\psi_{ii} = \text{Var}(U_i)$:

$$\begin{aligned} \Psi &= \frac{1}{n-1} \mathbf{U}' \mathbf{U} = \begin{pmatrix} \text{Var}(U_1) & \text{Cov}(U_1, U_2) & \cdots & \text{Cov}(U_1, U_m) \\ \text{Cov}(U_2, U_1) & \text{Var}(U_2) & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(U_m, U_1) & \text{Cov}(U_m, U_2) & \cdots & \text{Var}(U_m) \end{pmatrix} \\ &= \begin{pmatrix} \psi_{1,1} & 0 & \cdots & 0 \\ 0 & \psi_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \psi_{m,m} \end{pmatrix} \end{aligned}$$

Main Theorem of Factor Analysis & Program of Work

Main Theorem of Factor Analysis: The correlation matrix \mathbf{R} of the measured variables X_1, X_2, \dots, X_m satisfies the equation:

$$\mathbf{R} = \mathbf{A} \mathbf{A}' + \Psi$$

Program of Work:

- 1 **Principal Component Analysis for \mathbf{R} :** We ignore Ψ and solve $\mathbf{R} = \mathbf{A} \mathbf{A}'$ to determine/approximate the factor loading matrix \mathbf{A} .

or

Principal Component Analysis for \mathbf{R}_h : We estimate Ψ and then we solve $\mathbf{R}_h = \mathbf{A} \mathbf{A}'$ for the reduced correlation matrix $\mathbf{R}_h = \mathbf{R} - \Psi$ and determine/approximate the factor loading matrix \mathbf{A} .

- 2 Finally we determine the values of the factors with the help of the model equation $\mathbf{Z} = \mathbf{F} \mathbf{A}' + \mathbf{U}$.

Reduced Correlation Matrix and Communalities

As the covariance matrix Ψ of the unique factors U_1, U_2, \dots, U_m is diagonal, the reduced correlation matrix

$$\mathbf{R}_h = \mathbf{R} - \Psi$$

and \mathbf{R} are identical apart from the diagonal entries.

Communalities: The diagonal entries r_{jj} of \mathbf{R} are (see next slide):

$$r_{jj} = \text{Var}(Z_j) = \underbrace{\sum_{k=1}^p a_{jk}^2}_{= h_j^2} + \psi_{jj} = h_j^2 + \psi_{jj} \quad \Leftrightarrow \quad r_{jj} - \psi_{jj} = h_j^2$$

The number $h_j^2 = \sum_{k=1}^p a_{jk}^2$ is called the **communality** of Z_j .

- h_j^2 describes the part of the variance of Z_j that has been explained by the factors F_1, F_2, \dots, F_p .
- $h_j^2 = j$ th diagonal entry of $\mathbf{R}_h = \mathbf{R} - \Psi$.

Derivation of the Communalities

From $\mathbf{R} = \mathbf{A}\mathbf{A}' + \mathbf{\Psi}$ we obtain for the diagonal entries of \mathbf{R}

$$\begin{aligned}1 &= r_{jj} = \text{Cov}(Z_j, Z_j) = \text{Var}(Z_j) = \text{E} \left(\left[\sum_{k=1}^p a_{jk} \cdot F_k + U_j \right]^2 \right) \\&= \text{E} \left(\sum_{i=1}^p \sum_{k=1}^p a_{ji} \cdot a_{jk} \cdot F_i \cdot F_k + 2 \sum_{k=1}^p a_{jk} \cdot F_k \cdot U_j + U_j^2 \right) \\&= \sum_{i=1}^p \sum_{k=1}^p a_{ji} \cdot a_{jk} \cdot \text{E}(F_i \cdot F_k) + 2 \sum_{k=1}^p a_{jk} \cdot \text{E}(F_k \cdot U_j) + \text{E}(U_j^2) \\&= \sum_{i=1}^p \sum_{k=1}^p a_{ji} \cdot a_{jk} \cdot \underbrace{\text{Cov}(F_i, F_k)}_{\substack{= 0 \text{ if } i \neq k \\ = 1 \text{ if } i = k}} + 2 \sum_{k=1}^k a_{jk} \cdot \underbrace{\text{Cov}(F_k, U_j)}_{= 0} + \underbrace{\text{Var}(U_j)}_{= \psi_{jj}} \\&= \underbrace{\sum_{k=1}^p a_{jk}^2}_{= h_j^2} + \psi_{jj}, \quad \text{since } F_i, F_j \text{ are standardized and } U_j \text{ is centered.}\end{aligned}$$

Communalities Must be Estimated

As the a_{jk} are yet unknown, the communalities h_j^2 must be estimated with the help of the correlation matrix $\mathbf{R} = (r_{ij})$.

- 1 **Maximal absolute correlation (coefficient) of Z_j with all other variables:** Estimate the communality h_j^2 by

$$\hat{h}_j^2 = \max_{\substack{k=1, \dots, m, \\ k \neq j}} |r_{jk}|.$$

- 2 **Multiple determination coefficient:** Compute the inverse matrix $\mathbf{R}^{-1} = (q_{ij})$ of the correlation matrix \mathbf{R} and estimate h_j^2 by

$$\hat{h}_j^2 = 1 - \frac{1}{q_{jj}},$$

where q_{jj} is the j th diagonal entry of \mathbf{R}^{-1} .

- 3 An iterative scheme (not discussed here).

Ex. 2.4 (Communalities and Reduced Correlation Matrix): For our toy example, estimate the communalities with method 1 and estimate \mathbf{R}_h .

Computation of the Factor Loading Matrix with PCA I

- 1 Principal Component Analysis for solving $\mathbf{R} = \mathbf{A} \mathbf{A}'$
- 2 Principal Component Analysis for solving $\widehat{\mathbf{R}}_h = \mathbf{A} \mathbf{A}'$ with $\widehat{\mathbf{R}}_h =$ estimator of the reduced correlation matrix \mathbf{R}_h

\mathbf{R} and $\widehat{\mathbf{R}}_h$ are both symmetric; and for both cases we use the same method, called principal component analysis (PCA), described below.

Step 1: Compute the m real eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ and m corresponding orthonormal eigenvectors $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m$ of \mathbf{R} (or $\widehat{\mathbf{R}}_h$).

More precisely, we have $\mathbf{R} \mathbf{b}_j = \lambda_j \mathbf{b}_j$ (or $\widehat{\mathbf{R}}_h \mathbf{b}_j = \lambda_j \mathbf{b}_j$) and $\|\mathbf{b}_j\|_2 = 1$ for $j = 1, 2, \dots, m$, and $\mathbf{b}_j' \mathbf{b}_k = 0$ if $j \neq k$.

Step 2 (Initial Choice of \mathbf{A}): Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q > 0$ (positive eigenvalues), and let $0 \geq \lambda_{q+1} \geq \dots \geq \lambda_m$ (non-positive eigenvalues). Then choose \mathbf{A} preliminarily as:

$$\mathbf{A} = \left(\sqrt{\lambda_1} \mathbf{b}_1, \sqrt{\lambda_2} \mathbf{b}_2, \dots, \sqrt{\lambda_q} \mathbf{b}_q \right) \quad (m \times q \text{ matrix})$$

Step 3 (Reduction of \mathbf{A} to Include Only Significant Factors):

$$\mathbf{A} = \left(\sqrt{\lambda_1} \mathbf{b}_1, \sqrt{\lambda_2} \mathbf{b}_2, \dots, \sqrt{\lambda_p} \mathbf{b}_p \right) \quad (m \times p \text{ matrix})$$

where $p \leq q$ is determined as follows (different methods possible):

- **Cumulative Variance Criterion:** Determine p as the smallest integer such that at least $W\%$ of the variance are explained through the factors F_1, F_2, \dots, F_p , i.e.

$$\text{Var}(Z_j) = \sum_{k=1}^p a_{jk}^2 \geq \frac{W}{100} \quad \text{for } j = 1, \dots, m.$$

- **Kaiser Criterion:** $\lambda_1 \geq \dots \geq \lambda_p > 1$ and $1 \geq \lambda_{p+1} \geq \dots \geq \lambda_m$
- **Scree Test:** Plot (j, λ_j) , $j = 1, 2, \dots, m$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$, and connect the dots. Let $j = p$ be the largest integer before the curve has a sharp bend.

Ex. 2.5: Estimating the Factor Loading Matrix with PCA

Estimate the **factor loading matrix \mathbf{A}** for our toy example using the **Kaiser criterion**. Write down the explicit model equations and **interpret** them.

Excursion: Derivation of the Formula for \mathbf{A} – Part I

From the **eigenvalue-eigenvector relationships** $\mathbf{R} \mathbf{b}_j = \lambda_j \mathbf{b}_j, j = 1, 2, \dots, m,$

$$\mathbf{R} \underbrace{(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m)}_{=\mathbf{B}} = \underbrace{(\lambda_1 \mathbf{b}_1, \lambda_2 \mathbf{b}_2, \dots, \lambda_m \mathbf{b}_m)}_{=\mathbf{B}\mathbf{D}},$$

where $\mathbf{D} = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ is the diagonal matrix with the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m$ on the diagonal and zeros everywhere else.

As the matrix \mathbf{B} with the eigenvectors as columns is **orthogonal** (because the eigenvectors are orthonormal), multiplying with \mathbf{B}' from the right yields

$$\mathbf{R}\mathbf{B} = \mathbf{B}\mathbf{D} \quad \Leftrightarrow \quad \mathbf{R} = \mathbf{R} \underbrace{\mathbf{B}\mathbf{B}'}_{=\mathbf{I}} = \mathbf{B}\mathbf{D}\mathbf{B}'.$$

We now **omit all negative eigenvalues** and only use the positive eigenvalues $\lambda_1 \geq \dots \geq \lambda_q > 0$:

$$\mathbf{R} \approx \mathbf{B}\tilde{\mathbf{D}}\mathbf{B}' \quad \text{with} \quad \tilde{\mathbf{D}} = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_q, \underbrace{0, \dots, 0}_{m-q}) \quad (12)$$

Excursion: Derivation of the Formula for \mathbf{A} – Part II

As the non-zero diagonal entries in $\tilde{\mathbf{D}} = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_q, 0, \dots, 0)$ are all positive, we can **draw the root**

$$\tilde{\mathbf{D}} = \tilde{\mathbf{D}}^{1/2} \tilde{\mathbf{D}}^{1/2} \text{ with } \tilde{\mathbf{D}}^{1/2} = \text{Diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_q}, \underbrace{0, \dots, 0}_{m-q}) = (\tilde{\mathbf{D}}^{1/2})'$$

Substituting $\tilde{\mathbf{D}} = \tilde{\mathbf{D}}^{1/2} \tilde{\mathbf{D}}^{1/2}$ in (12), we find

$$\mathbf{R} \approx \mathbf{B} \tilde{\mathbf{D}} \mathbf{B}' = \mathbf{B} \tilde{\mathbf{D}}^{1/2} \underbrace{\tilde{\mathbf{D}}^{1/2} \mathbf{B}'}_{=(\tilde{\mathbf{D}}^{1/2})'} = \underbrace{(\mathbf{B} \tilde{\mathbf{D}}^{1/2})}_{=\tilde{\mathbf{A}}} \underbrace{(\tilde{\mathbf{D}}^{1/2} \mathbf{B}')'}_{=\tilde{\mathbf{A}}} = \tilde{\mathbf{A}} \tilde{\mathbf{A}}'$$

We compute $\tilde{\mathbf{A}} = \mathbf{B} \tilde{\mathbf{D}}^{1/2} = (\sqrt{\lambda_1} \mathbf{b}_1, \dots, \sqrt{\lambda_q} \mathbf{b}_q, \mathbf{0}_{\mathbf{b}_{q+1}}, \dots, \mathbf{0}_{\mathbf{b}_m})$

$$= (\underbrace{\sqrt{\lambda_1} \mathbf{b}_1, \dots, \sqrt{\lambda_q} \mathbf{b}_q}_{m \times q}, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{m \times (m-q)})$$

and 'simplify' $\tilde{\mathbf{A}}$ by dropping the last $m - q$ columns that are zero vectors:

$$\mathbf{A} = (\sqrt{\lambda_1} \mathbf{b}_1, \dots, \sqrt{\lambda_q} \mathbf{b}_q) \quad (m \times q \text{ matrix}).$$

Computation of the Values of the Factors

Model:
$$\underbrace{\mathbf{Z}'}_{m \times n} = \underbrace{\mathbf{A}}_{m \times p} \underbrace{\mathbf{F}'}_{p \times n} + \underbrace{\mathbf{U}'}_{m \times n} \Leftrightarrow \mathbf{Z} = \mathbf{F} \mathbf{A}' + \mathbf{U}$$

- \mathbf{U} contains the effects not explained through the linear model $\mathbf{F} \mathbf{A}'$. We want the entries of \mathbf{U} to be as small as possible.
 - For $p < m$ the equation $\mathbf{Z}' = \mathbf{A} \mathbf{F}'$ does usually not have a solution.
-

Least Squares Approach: We try to find \mathbf{F} such that $\mathbf{Z}' - \mathbf{A} \mathbf{F}'$ is as close to the zero matrix as possible. Multiply $\mathbf{Z}' = \mathbf{A} \mathbf{F}'$ from the left by \mathbf{A}'

$$\mathbf{A}' \mathbf{Z}' = \underbrace{\mathbf{A}' \mathbf{A}}_{p \times p} \mathbf{F}'$$

where we know that $\mathbf{A}' \mathbf{A}$ is invertible because $\text{rank}(\mathbf{A}) = p$. Hence

$$\mathbf{F}' = (\mathbf{A}' \mathbf{A})^{-1} \mathbf{A}' \mathbf{Z}'$$

Ex. 2.6 (Factor Values): Compute the factor values for our toy example.

Interpretation of the Values of the Factors

For the k th object e_k , we have a vector $(f_{k1}, f_{k2}, \dots, f_{kp})$ of factor values:

standardized j th random variable

Z_j measured for k th object e_k :

$$z_{kj} = a_{j1} \cdot f_{k1} + a_{j2} \cdot f_{k2} + \dots + a_{jp} \cdot f_{kp}$$

By assumption, the factors are standardized: $E(F_j) = 0$ and $\text{Var}(F_j) = 1$.

Interpretation of the factor values of object k :

- If $f_{ki} = 0$ then the effect of factor F_i on object e_k is **average**.
 - If $f_{ki} > 0$ then the effect of factor F_i on object e_k is **above average**.
 - If $f_{ki} < 0$ then the effect of factor F_i on object e_k is **below average**.
-

Ex. 2.7 (Interpretation): Interpret the factor values for our example.

Outlook: Method of Factor Rotation

We have constructed a $m \times p$ factor loading matrix \mathbf{A} such that

$$\mathbf{R} = \mathbf{A}\mathbf{A}' + \hat{\Psi},$$

where $\hat{\Psi}$ is an approximation of the diagonal correlation matrix Ψ of the unique factors (note that usually the approximation $\hat{\Psi}$ will not be diagonal).

If we replace \mathbf{A} by $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{T}$, where \mathbf{T} is an $p \times p$ rotation matrix (i.e. $\det(\mathbf{T}) = 1$ and $\mathbf{T}'\mathbf{T} = \mathbf{T}\mathbf{T}' = \mathbf{I}$) then

$$\tilde{\mathbf{A}}\tilde{\mathbf{A}}' + \hat{\Psi} = (\mathbf{A}\mathbf{T})(\mathbf{A}\mathbf{T})' + \hat{\Psi} = \mathbf{A}\underbrace{\mathbf{T}\mathbf{T}'}_{=\mathbf{I}}\mathbf{A} + \hat{\Psi} = \mathbf{A}\mathbf{A}' + \hat{\Psi} = \mathbf{R}$$

Clearly, $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{T}$ is also a $m \times p$ factor loading matrix. (The values of the factors have to be recomputed for this new factor loading matrix).

Outlook: Why is Factor Rotation Useful?

Why can a suitable rotation of the factor loading matrix be useful?

The entries of the factor loading matrix are the coefficients in

$$Z_j = a_{j1} \cdot F_1 + a_{j2} \cdot F_2 + \dots + a_{jp} \cdot F_p + U_j, \quad j = 1, 2, \dots, m.$$

- These equations are best interpreted if the a_{ij} are either close to zero or close to one.
- By our construction we are likely to get many coefficients satisfying $0.3 \leq |a_{ij}| \leq 0.7$ and these are difficult to interpret.
- By choosing a suitable rotation matrix \mathbf{T} we may obtain more coefficients in the new factor loading matrix $\tilde{\mathbf{A}} = \mathbf{A} \mathbf{T}$ that are either close to zero or close to one and thus are better to interpret.

Structural Equation Modeling

Topic 3: Introduction to Structural Equation Modeling (SEM)

Dr. Kerstin Hesse

Email: kerstin.hesse@hhl.de; *Phone:* +49 (0)341 9851-820; *Office:* HHL Main Building, Room 115A

HHL Leipzig Graduate School of Management, Jahnallee 59, 04109 Leipzig, Germany

Doctoral Program at HHL, June 1-2, 2012

Topic 3: Introduction to Structural Equation Modeling

- Structural (Inner) Model
 - causal (cause-effect) relationships
 - exogenous and endogenous latent variables
 - regression equations of the structural (inner) model
- (Outer) Measurement Models
 - formative measurement model/block
 - reflective measurement model/block
- Approaches to Structural Equation Modeling
 - LISREL (linear structural relationships) – discussed in detail as Topic 4
 - PLS (partial least squares) path modeling – discussed in detail as Topic 5

Aim of Structural Equation Modeling: Causal Modeling

Aim: Structural equation modeling (SEM) aims at describing **causal relationships** of **latent (non-measurable) variables** in a **linear model**.

Causal Relationship: A causal relationship is a **cause-effect relationship**. Depicted graphically the **arrow points from the cause to the effect**.

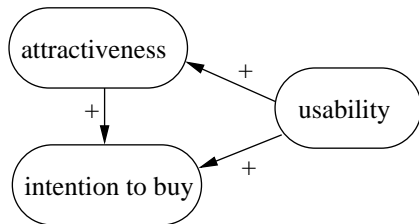
Latent Variables: A latent variable is a (dependent or independent) variable in a model that **(usually) cannot be directly measured**.

Example (Causal Relationships for Margarine):

The more widely a margarine can be used, the more attractive it is.

The more attractive a margarine is, the more likely it is to be bought.

The more widely a margarine can be used, the more likely it is to be bought.



Exogenous and Endogenous Latent Variables

Example (Causal Relationships for Margarine) Continued:

The **latent variables** are: attractiveness, usability and intention to buy.

The name **latent** refers to the fact that they are variables for 'abstract concepts' that (usually) cannot be directly measured.

The **plus signs at the arrows** indicate that we have a positive correlation. If we would replace 'intention to buy' by 'disinclination to buy', then we would get **negative signs** on the two arrows pointing to that latent variable (e.g. 'The higher the attractiveness the lower the disinclination to buy.').

Exogenous and Endogenous Latent Variables:

Exogenous latent variables are **independent** latent variables.

Endogenous latent variables are **dependent** latent variables.

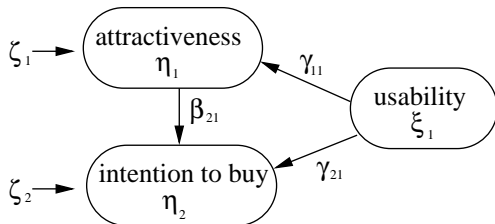
Example Continued: Usability is an exogenous latent variable.
Attractiveness and intention to buy are endogenous latent variables.

Example: Modeling the Structural (Inner) Model I

Assumptions: All latent variables can be represented on a **metric scale**.

Model: Regression type equations model the causal relationships.

Example (Causal Relationships for Margarine) Continued:



Linear model equations:

$$\eta_1 = \gamma_{1,1} \xi_1 + \zeta_1$$

$$\eta_2 = \beta_{2,1} \eta_1 + \gamma_{2,1} \xi_1 + \zeta_2$$

where ζ_1, ζ_2 are **error terms**.

Rules for denoting the coefficients $\gamma_{1,1}$, $\gamma_{2,1}$ and $\beta_{2,1}$:

γ is used for arrows pointing from an exogenous to an endogenous latent variable. β is used for arrows pointing from an endogenous to another endogenous latent variable. **Indices:** The first index is the one of the affected variable and the second index is the one of the causing variable.

Example: Modeling the Structural (Inner) Model II

Matrix Equations: We can also write our model in **matrix form**:

$$\underbrace{\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}}_{= \boldsymbol{\eta}} = \underbrace{\begin{pmatrix} 0 & 0 \\ \beta_{2,1} & 0 \end{pmatrix}}_{= \mathbf{B}} \underbrace{\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}}_{= \boldsymbol{\eta}} + \underbrace{\begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \end{pmatrix}}_{= \boldsymbol{\Gamma}} \xi_1 + \underbrace{\begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}}_{= \boldsymbol{\zeta}}$$

and shorter

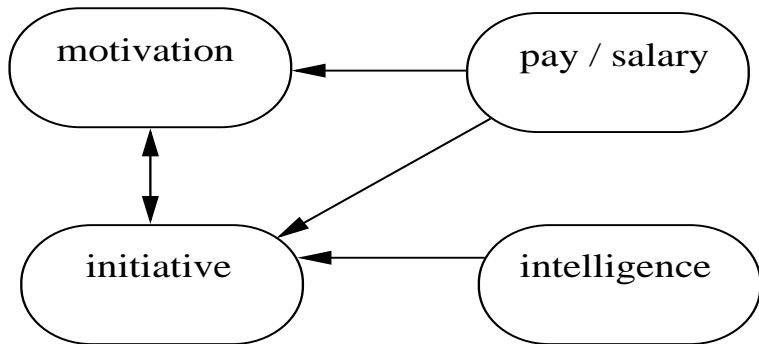
$$\boldsymbol{\eta} = \mathbf{B} \boldsymbol{\eta} + \boldsymbol{\Gamma} \boldsymbol{\xi} + \boldsymbol{\zeta},$$

where (notation **not** restricted to this example):

- $\boldsymbol{\xi}$ = vector of exogenous latent variables
- $\boldsymbol{\eta}$ = vector of endogenous latent variables
- $\boldsymbol{\zeta}$ = vector of error terms for the endogenous latent variables
- \mathbf{B} = coefficients for the relationships among the endogenous latent variables (\mathbf{B} has always zeros on the diagonal.)
- $\boldsymbol{\Gamma}$ = coefficients for the effects of the exogenous latent variables on the endogenous latent variables

Ex. 3.1: Setting up the Structural (Inner) Model

A model for the work of a software programmer on a non-pay-scale salary is shown in the diagram below. Indicate the various latent variables and the coefficients and error terms in the diagram, using the rules explained on pages 71–72. For consistency, number any exogenous (or endogenous) latent variables from top to bottom. Finally write down the equations for the structural (inner) model.



Summary: Setting up the Structural (Inner) Model

Random Variables:

$\xi = (\xi_1, \dots, \xi_s)'$ = vector of **exogenous (independent)** latent variables ξ_i

$\eta = (\eta_1, \dots, \eta_r)'$ = vector of **endogenous (dependent)** latent variables η_i

$\zeta = (\zeta_1, \dots, \zeta_r)'$ = vector of **error terms** ζ_i in regression equations for η_i

Linear Regression Equations: The latent variables are related by

$$\eta_i = \beta_{i1} \eta_1 + \beta_{i2} \eta_2 + \dots + \beta_{ir} \eta_r + \gamma_{i1} \xi_1 + \gamma_{i2} \xi_2 + \dots + \gamma_{is} \xi_s + \zeta_i.$$

(Path) Coefficients in Regression Equations:

$\Gamma = (\gamma_{ij}) = r \times s$ matrix with the coefficients γ_{ij} for the effect of ξ_j on η_i

$\mathbf{B} = (\beta_{ij}) = r \times r$ matrix with the coefficients β_{ij} for the effect of η_j on η_i , where $\beta_{ii} = 0$ for all $i = 1, 2, \dots, r$

Structural (Inner) Model in Matrix Form:

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

Reduced Model and Graphic Representation

Reduced Model: We collect all terms involving η on the left-hand side and solve for η :

$$\begin{aligned}\eta &= \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta & \Leftrightarrow & \quad \eta - \mathbf{B}\eta = \mathbf{\Gamma}\xi + \zeta & \Leftrightarrow \\ \underbrace{(\mathbf{I} - \mathbf{B})}_{= \mathbf{B}^*} \eta &= \mathbf{\Gamma}\xi + \zeta & \Leftrightarrow & \quad \eta = (\mathbf{B}^*)^{-1} \mathbf{\Gamma}\xi + (\mathbf{B}^*)^{-1} \zeta\end{aligned}$$

Ex. 3.2 (Reduced Model): Write down the **reduced model** for the structural (inner) model from Ex. 3.1.

Graphic Representation of the Structural (Inner) Model: It helps to draw a diagram of the structural model as done in the examples. **Rules:**

- Arrows point from cause to effect.
- Latent variables are in circles or ellipsoids (here: rounded boxes)
- An arrow points from errors ζ_i (no box) to the corresponding η_i .

Assumptions in the Structural (Inner) Model

$$\eta_i = \beta_{i1} \eta_1 + \beta_{i2} \eta_2 + \dots + \beta_{ir} \eta_r + \gamma_{i1} \xi_1 + \gamma_{i2} \xi_2 + \dots + \gamma_{is} \xi_s + \zeta_i.$$

The latent variables are **metric but have no fixed zero point**. We choose

$$E(\xi_i) = 0 \quad \text{and} \quad E(\eta_i) = 0 \quad \text{for all } i,$$

and this implies automatically for the error terms $E(\zeta_i) = 0$ for all i .

For LISREL (but not necessarily for PLS path modeling), we assume that the independent latent variables ξ_i and the errors ζ_j are **uncorrelated**:

$$\text{Cov}(\xi_i, \zeta_j) = E(\xi_i \zeta_j) = 0 \quad \text{for all } i, j.$$

The error terms ζ_i **may correlate with each other**:

$$\psi_{ij} = \text{Cov}(\zeta_i, \zeta_j) \quad \text{and} \quad \psi_{ii} = \text{Cov}(\zeta_i, \zeta_i) = \text{Var}(\zeta_i)$$

Comments on the Structural (Inner) Model

The exogenous latent variables ξ_i may correlate with each other:

$$\phi_{ij} = \text{Cov}(\xi_i, \xi_j) \quad \text{and} \quad \phi_{ii} = \text{Cov}(\xi_i, \xi_i) = \text{Var}(\xi_i)$$

$$\begin{aligned} \eta &= \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta & \Leftrightarrow & \quad \eta - \mathbf{B}\eta = \mathbf{\Gamma}\xi + \zeta & \Leftrightarrow \\ \underbrace{(\mathbf{I} - \mathbf{B})}_{= \mathbf{B}^*} \eta &= \mathbf{\Gamma}\xi + \zeta & \Leftrightarrow & \quad \eta = (\mathbf{B}^*)^{-1} \mathbf{\Gamma}\xi + (\mathbf{B}^*)^{-1} \zeta \end{aligned}$$

The matrix \mathbf{B}^* will be invertible if we have no redundant equations.

The assumptions on the previous page can also be written in vector notation:

$$\mathbf{E}(\xi) = \mathbf{0}, \quad \mathbf{E}(\eta) = \mathbf{0}, \quad \mathbf{E}(\zeta) = \mathbf{0},$$

and for LISREL (but not necessarily for PLS): $\text{Cov}(\xi, \zeta) = \mathbf{0}$,

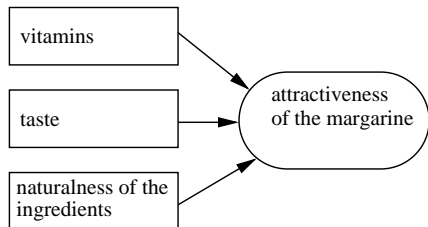
(Outer) Measurement Models for the Latent Variables

Latent variables (usually) cannot be measured directly, but we assume that they can be described on a metric scale.

Therefore we need measurement variables (indicator variables) that are directly linked to the latent variables and that can be measured. We have two types of measurement models:

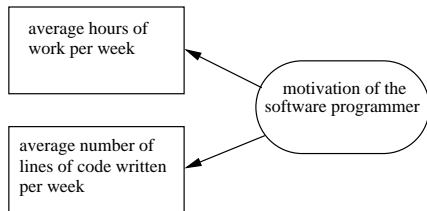
Formative Measurement Model:

The measured variables influence the latent variable.



Reflective Measurement Model:

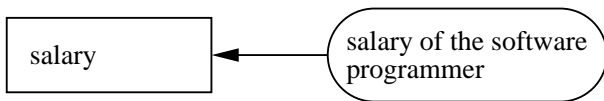
The latent variable influences the measured variables.



Rules for Measurement Models

Before we can set up the equations for the measurement models we need some **notation conventions/rules**:

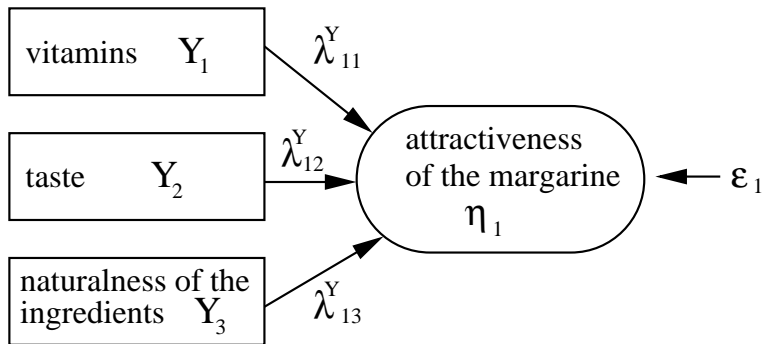
- Measurement variables are depicted in **angular boxes**.
- Measurement variables for an **exogenous** latent variable are called X_i .
- Measurement variables for an **endogenous** latent variable are called Y_i .
- If a latent variable **can be measured directly** then we may take the latent variable **itself** as its own measurement variable.



Actually here the direction of the error does not really matter as 'salary = salary of the software programmer'.

- Non-measurable latent variables should usually have **at least two or better three** measurement variables.

Example: Formative Measurement Model

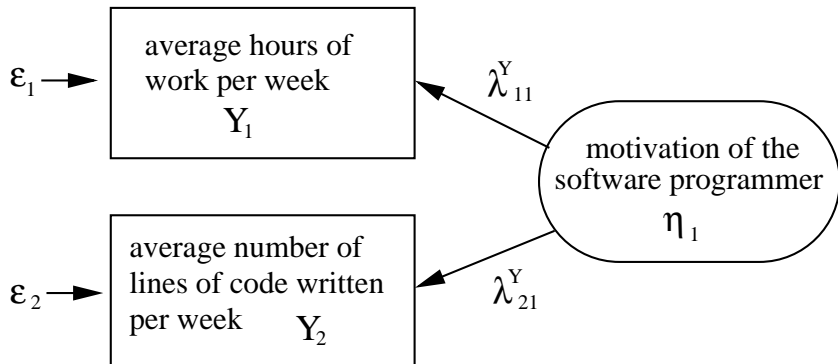


Formative Measurement Model with Regression Equation:

$$\eta_1 = \lambda_{1,1}^Y (Y_1 - \mu_{Y_1}) + \lambda_{1,2}^Y (Y_2 - \mu_{Y_2}) + \lambda_{1,3}^Y (Y_3 - \mu_{Y_3}) + \varepsilon_1,$$

where μ_{Y_1} , μ_{Y_2} and μ_{Y_3} are the expectation values of Y_1 , Y_2 and Y_3 ; $\lambda_{1,1}^Y$, $\lambda_{1,2}^Y$, $\lambda_{1,3}^Y$ are the coefficients; and ε_1 is the error term.

Example: Reflective Measurement Model



Reflective Measurement Model with Factor Analytic Equations:

$$Y_1 - \mu_{Y_1} = \lambda_{1,1}^Y \eta_1 + \epsilon_1,$$

$$Y_2 - \mu_{Y_2} = \lambda_{2,1}^Y \eta_1 + \epsilon_2,$$

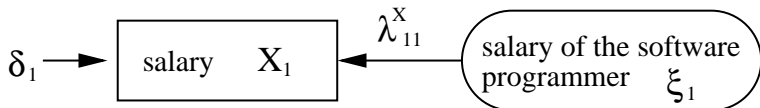
where μ_{Y_1} and μ_{Y_2} are the expectation values of Y_1 and Y_2 ; $\lambda_{1,1}^Y$, $\lambda_{2,1}^Y$ are the coefficients; and ϵ_1 , ϵ_2 are the error terms.

More on Measurement Models

Note: So far we saw measurement models for **endogenous** latent variables. Measurement models for **exogenous latent variables** look just the same, with two **notational differences**:

- Measurement variables are called X_1, X_2, \dots (instead of Y_1, Y_2, \dots).
- Coefficients get an upper index X instead of Y .
- Error terms are called $\delta_1, \delta_2, \dots$ (instead of $\varepsilon_1, \varepsilon_2, \dots$).

Latent Variables that can be Measured Directly Without Error:



A reflective measurement model (for the latent exogenous variable ξ_1) is

$$X_1 - \mu_{X_1} = \lambda_{1,1}^X \xi_1 + \delta_1 = \xi_1 \quad \text{with} \quad \lambda_{1,1}^X = 1 \quad \text{and} \quad \delta_1 = 0,$$

since the (centered) salary $X_1 - \mu_{X_1}$ equals the latent variable ξ_1 and since we assume that it can be measured without error.

Summary: Setting up a Formative Measurement Model

Linear Regression Equation for an Endogenous Latent Variable η_i :

$$\eta_i = \lambda_{i1}^Y (Y_1 - \mu_{Y_1}) + \lambda_{i2}^Y (Y_2 - \mu_{Y_2}) + \dots + \lambda_{ip}^Y (Y_p - \mu_{Y_p}) + \varepsilon_i,$$

where: Y_1, Y_2, \dots, Y_p are the measurement variables related to η_i ;

$\mu_{Y_1}, \mu_{Y_2}, \dots, \mu_{Y_p}$ are the expectation values of Y_1, Y_2, \dots, Y_p ;

λ_{ij}^Y = coefficient for the effect of $(Y_j - \mu_{Y_j})$ on η_i ; ε_i = error term for η_i .

Linear Regression Equation for an Exogenous Latent Variable ξ_i :

$$\xi_i = \lambda_{i1}^X (X_1 - \mu_{X_1}) + \lambda_{i2}^X (X_2 - \mu_{X_2}) + \dots + \lambda_{iq}^X (X_q - \mu_{X_q}) + \delta_i,$$

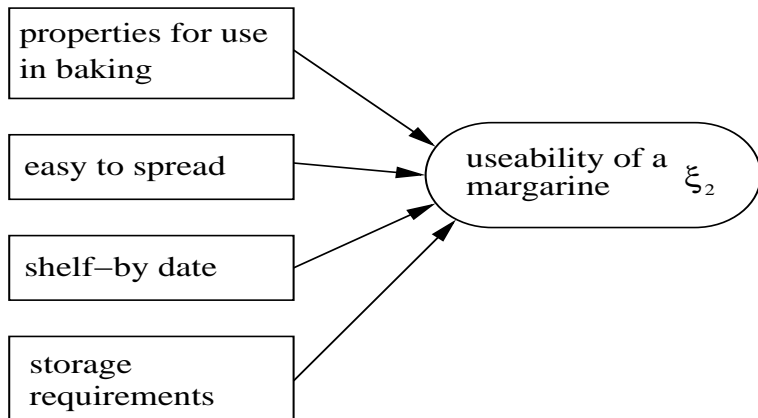
where: X_1, X_2, \dots, X_q are the measurement variables related to ξ_i ;

$\mu_{X_1}, \mu_{X_2}, \dots, \mu_{X_q}$ are the expectation values of X_1, X_2, \dots, X_q ;

λ_{ij}^X = coefficient for the effect of $(X_j - \mu_{X_j})$ on ξ_i ; δ_i = error term for ξ_i .

Ex. 3.3: Formative Measurement Model

Starting the numbering of the measurement variables at the top, indicate the **measurement variables**, **error terms** and **coefficients** in the following diagram of a **formative measurement model**. Then write down the **regression equation** for the exogenous latent variable ξ_2 .



Summary: Setting up a Reflective Measurement Model

Factor Analytic Equations for Measuring Endogenous Latent Variables η_j :

$$Y_i - \mu_{Y_i} = \lambda_{i1}^Y \eta_1 + \lambda_{i2}^Y \eta_2 + \dots + \lambda_{ir}^Y \eta_r + \varepsilon_i \quad (13)$$

where: $\mu_{Y_i} = E(Y_i)$; η_1, \dots, η_r are the endogenous latent variables; λ_{ij}^Y is the coefficient for the effect of η_j on Y_i ; ε_i is the error of Y_i .

Factor Analytic Equations for Measuring Exogenous Latent Variables ξ_j :

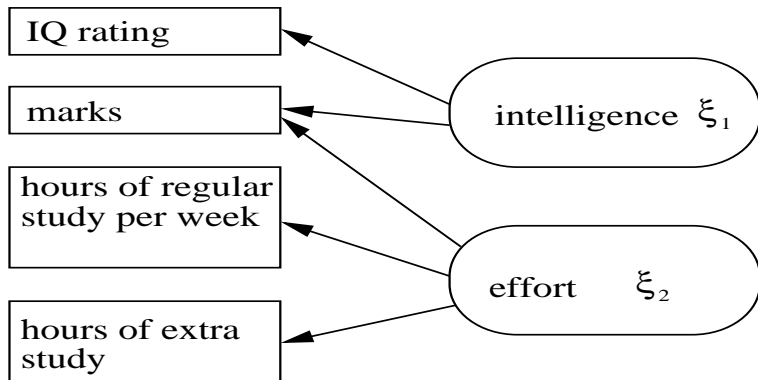
$$X_i - \mu_{X_i} = \lambda_{i1}^X \xi_1 + \lambda_{i2}^X \xi_2 + \dots + \lambda_{is}^X \xi_s + \delta_i \quad (14)$$

where: $\mu_{X_i} = E(X_i)$; ξ_1, \dots, ξ_s are the exogenous latent variables; λ_{ij}^X is the coefficient for the effect of ξ_j on X_i ; δ_i is the error of X_i .

In (13) and (14) usually **only a few (often only one) coefficients** are $\neq 0$.

Ex. 3.4: Reflective Measurement Model

The diagram below shows part of a structural equation model for the academic success of students. Numbering the measurement variables from the top to the bottom, complete the diagram of the reflective measurement model by indicating the **variables**, **error terms** and **coefficients**. Then write down the **factor analytic equations** for the measurement variables.



Preparations for Structural Equation Modeling (SEM)

The structural model needs to be **based on logical cause-effect relationships** that are supported by evidence!

Structural equation modeling **cannot prove** a model, but it can **only give empirical support** to a model.

It is crucial to determine the **correct cause-effect relationships**:

- Which variables have formative measurement models?
- Which variables have reflective measurement models?

If a structural equation model is true, then causal relationships should be reflected by **non-zero covariances** between related measurement variables.

Attention: Covariance can only measure **linear** relationships!

A non-zero covariance can be **coincidental**, rather than reflect a relationship!

A non-zero covariance can be **caused by a third variable** on which the correlated variables depend and that is **not even included in the model!**

Two Approaches to SEM: LISREL and PLS Path Modeling

LISREL (Linear Structural Relationships):

- is based on approximating the **covariance matrix** of the measured variables
- usually requires **reflective** measurement models

PLS (Partial Least Squares) Path Modeling:

- aims to **maximize the variance** explained in the model
- can work with both **formative and reflective** measurement models or a mixture of both

Other Differences Between the Models:

- LISREL assumes a **multivariate normal distribution** and allows a good assessment of the quality of the model (if these distributional assumptions are satisfied).
- PLS path modeling makes less assumptions on the data, but also allows fewer information about the quality of the model.

Further Discussion of Structural Equation Modeling Will be Given Separately for LISREL and PLS Path Modeling

From here we will **discuss the two approaches separately**:

- The two approaches use a somewhat **different notation for the model**, i.e. for the combination of the structural (inner) model and the measurement model.
 - The two approaches seem to have **different strengths** and **different weaknesses**.
-

Literature (comments based on my own literature review):

- There is not much (statistical) literature on structural equation modeling (either approach).
- LISREL seems to have been statistically well developed, whereas there seems to be hardly any statistical literature on PLS path modeling.

Structural Equation Modeling

Topic 4: LISREL (Linear Structural Relationships)

Dr. Kerstin Hesse

Email: kerstin.hesse@hhl.de; *Phone:* +49 (0)341 9851-820; *Office:* HHL Main Building, Room 115A

HHL Leipzig Graduate School of Management, Jahnallee 59, 04109 Leipzig, Germany

Doctoral Program at HHL, June 1-2, 2012

Topic 4: LISREL (Linear Structural Relationships)

- **Setting Up the Structural Equation Model for LISREL**

- structural (inner) model in LISREL and assumptions
- reflective measurement models in LISREL (LISREL normally requires reflective measurement models/blocks!)

- **LISREL's Covariance Approach**

- (empirical) covariance matrix of the data of the measurement variables
- idea of LISREL: compare the empirical with the 'theoretical' covariance matrix to determine the model parameters
- example and exercise: solving a simple LISREL model by hand
- how LISREL models are solved in practice with the help of a computer

- **Appendix: LISREL in Matrix Notation**

As textbooks and research papers often write down LISREL in the (less intuitive) matrix notation, this has been included in an appendix.

General Observations on LISREL

General Assumption: The variables are assumed to follow a **multivariate normal distribution**.

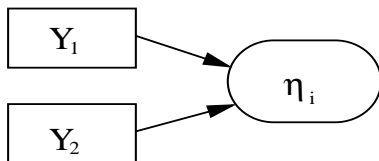
In the usual literature the measurement models are assumed to be **reflective**, although some references claim that LISREL can also work with **formative measurement models**.

Suggestion for Formative Measurement Models: One book made the suggestion that each measurement variable of a formative measurement model block could be included in the inner model as a **exogenous latent variable** that is then **measured directly by itself** with coefficient $\lambda^X = 1$. This is illustrated on the next slide.

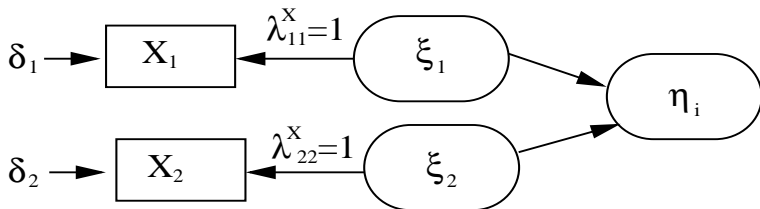
A drawback of this idea seems to be that this leads to an **increase of the number of exogenous latent variables** in the inner model.

Converting a Formative to a Reflective Measurement Block

Original formative measurement block for η_i :



New exogenous latent variables that are measured directly (reflective):



We have $X_1 = Y_1$, $X_2 = Y_2$ and $X_1 = \xi_1 + \delta_1$, $X_2 = \xi_2 + \delta_2$, where the errors δ_1 and δ_2 are only $\neq 0$ if X_1 and X_2 are not measured exactly.

Setting up Reflective Measurement Models in LISREL I

Measurement Variables of Endogenous Latent Var. and their Errors:

The measurement variables Y_1, \dots, Y_p of the endogenous latent variables η_1, \dots, η_r are numbered consecutively (no relation to the numbering of the endogenous latent variables η_1, \dots, η_r).

$$\mathbf{y} = (Y_1, \dots, Y_p)' = \begin{pmatrix} \text{vector of all measurement variables for the} \\ \text{endogenous latent variables } \boldsymbol{\eta} = (\eta_1, \dots, \eta_r)' \end{pmatrix}$$
$$\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p)' = \begin{pmatrix} \text{vector of all measurement errors for the} \\ \text{measurement variables } \mathbf{y} = (Y_1, \dots, Y_p)' \end{pmatrix}$$

Factor Analytic Equation for an Endogenous Latent Variable:

$$Y_i - \mu_{Y_i} = \lambda_{i1}^Y \eta_1 + \lambda_{i2}^Y \eta_2 + \dots + \lambda_{ir}^Y \eta_r + \varepsilon_i,$$

where: λ_{ij}^Y is the coefficient for the effect of η_j on Y_i ; ε_i is the error of Y_i .

Measurement Variables of Exogenous Latent Var. and their Errors:

The measurement variables X_1, \dots, X_q of the exogenous latent variables ξ_1, \dots, ξ_s are numbered consecutively (no relation to the numbering of the endogenous latent variables ξ_1, \dots, ξ_s).

$$\mathbf{x} = (X_1, \dots, X_q)' = \begin{pmatrix} \text{vector of all measurement variables for the} \\ \text{endogenous latent variables } \boldsymbol{\xi} = (\xi_1, \dots, \xi_s)' \end{pmatrix}$$
$$\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)' = \begin{pmatrix} \text{vector of all measurement errors for the} \\ \text{measurement variables } \mathbf{x} = (X_1, \dots, X_q)' \end{pmatrix}$$

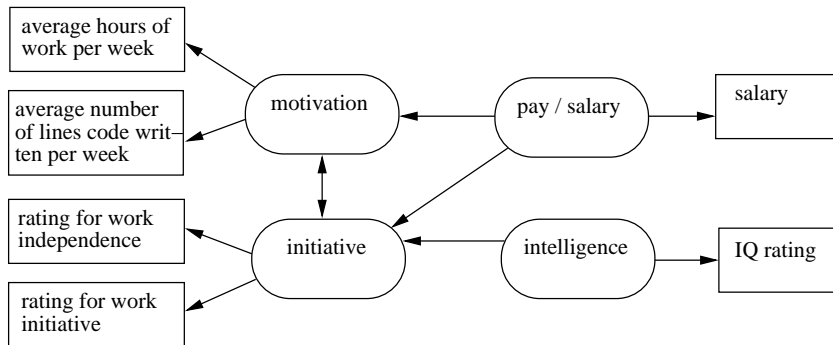
Factor Analytic Equation for an Exogenous Latent Variable:

$$X_i - \mu_{X_i} = \lambda_{i1}^X \xi_1 + \lambda_{i2}^X \xi_2 + \dots + \lambda_{is}^X \xi_s + \delta_i$$

where: λ_{ij}^X is the coefficient for the effect of ξ_j on X_i ; δ_i is the error of X_i .

Ex. 4.1: Structural Equation Model

The structural model for the **work of a software programmer on a non-pay-scale salary** (see Ex. 3.1) has now been equipped with the **reflective measurement models** for the latent variables shown below. Indicate all variables, errors and coefficients in the diagram and **write down the equations of the measurement models**. The ratings (apart from the IQ one) have been provided by the programmer's superior.



LISREL: Structural (Inner) Model with Assumptions

ξ_1, \dots, ξ_s are the exogenous (independent) latent variables,

η_1, \dots, η_r are the endogenous (dependent) latent variables,

ζ_i is the error term in regression equations for η_i , $i = 1, \dots, r$

γ_{ij} is the coefficient for the effect of ξ_j on η_i

β_{ij} is the coefficient for effect of η_j on η_i , where $\beta_{ii} = 0$ for $i = 1, 2, \dots, r$

Linear Regression Equations: The latent variables are related by

$$\eta_i = \beta_{i1} \eta_1 + \beta_{i2} \eta_2 + \dots + \beta_{ir} \eta_r + \gamma_{i1} \xi_1 + \gamma_{i2} \xi_2 + \dots + \gamma_{is} \xi_s + \zeta_i.$$

Assumptions: For all i and for all j

$$E(\xi_i) = 0, \quad E(\eta_i) = 0, \quad E(\zeta_i) = 0, \quad \text{Cov}(\xi_i, \zeta_j) = 0$$

Parameters: coefficients in regression equations γ_{ij} , β_{ij} and (co)variances $\phi_{ij} = \text{Cov}(\xi_i, \xi_j)$, $\phi_{ii} = \text{Var}(\xi_i)$, and $\psi_{ij} = \text{Cov}(\zeta_i, \zeta_j)$, $\psi_{ii} = \text{Var}(\zeta_i)$

Other Unknowns: values of latent variables, $\text{Cov}(\eta_i, \eta_j)$, $i, j = 1, \dots, r$

LISREL: Reflective Measurement Models I

Y_1, \dots, Y_p are the measurement variables for η_1, \dots, η_r ,

ε_i is the measurement error for Y_i , and $\mu_{Y_i} = E(Y_i)$.

λ_{ij}^Y is the coefficient for the effect of η_j on Y_i .

X_1, \dots, X_q are the measurement variables for ξ_1, \dots, ξ_s ,

δ_i is the measurement error for X_i , and $\mu_{X_i} = E(X_i)$.

λ_{ij}^X is the coefficient for the effect of ξ_j on X_i .

Factor Analytic Equations for the Measurement Variables:

$$Y_i - \mu_{Y_i} = \lambda_{i1}^Y \eta_1 + \lambda_{i2}^Y \eta_2 + \dots + \lambda_{ir}^Y \eta_r + \varepsilon_i$$

$$X_i - \mu_{X_i} = \lambda_{i1}^X \xi_1 + \lambda_{i2}^X \xi_2 + \dots + \lambda_{is}^X \xi_s + \delta_i$$

Assumptions:

- $E(\varepsilon_i) = 0$ and $E(\delta_i) = 0$ for all i , i.e. the measurement errors have all expectation value zero.

Assumptions (continued):

- The errors do **not correlate** with the latent variables: for all i and j
$$\text{Cov}(\varepsilon_i, \eta_j) = 0, \quad \text{Cov}(\varepsilon_i, \xi_j) = 0, \quad \text{Cov}(\delta_i, \eta_j) = 0, \quad \text{Cov}(\delta_i, \xi_j) = 0.$$
- The errors for the different types of measurement variables do **not correlate** with each other:

$$\text{Cov}(\varepsilon_i, \delta_j) = 0 \quad \text{for all } i \text{ and all } j.$$

- The errors ζ_j for the endogenous latent variables η_j do **not correlate** with the measurement errors ε_i and δ_i :

$$\text{Cov}(\varepsilon_i, \zeta_j) = 0 \quad \text{and} \quad \text{Cov}(\delta_i, \zeta_j) = 0 \quad \text{for all } i \text{ and all } j.$$

Parameters: coefficients λ_{ij}^X , λ_{ij}^Y in the measurement models and (co)variances $\theta_{ij}^\varepsilon = \text{Cov}(\varepsilon_i, \varepsilon_j)$, $\theta_{ii}^\varepsilon = \text{Var}(\varepsilon_i)$, and (co)variances $\theta_{ij}^\delta = \text{Cov}(\delta_i, \delta_j)$, $\theta_{ii}^\delta = \text{Var}(\delta_i)$

Idea: Compare Empirical with Model Covariance Matrix

The **only measurable quantities** in the LISREL approach are the **measurement variables (indicator variables)** X_i and Y_j .

From the structural equation model, the measurement variables are linked via the paths that connect latent variables and measurement variables.

Hence, we expect that the **measurement variables** $X_i, i = 1, 2, \dots, q$, and $Y_j, j = 1, 2, \dots, p$, are **correlated**.

Comparison of Empirical and Model Covariance Matrix:

From the data for X_1, \dots, X_q and Y_1, \dots, Y_p , we can compute the **empirical covariance matrix** of the **measurement variables**.

From the equations of our SEM, we can compute the **model covariance matrix** of the **measurement variables** in terms of the **model parameters**.

Equating the two allows us to estimate the model parameters of the SEM.

Data Matrix & Estimating the Expectation Values

Data Matrix: Let e_1, e_2, \dots, e_n be the objects on which we have measured the measurement variables $X_1, \dots, X_q, Y_1, \dots, Y_p$. The **data matrix** is

$$(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} x_{1,1} & \cdots & x_{1,q} & y_{1,1} & \cdots & y_{1,p} \\ x_{2,1} & \cdots & x_{2,q} & y_{2,1} & \cdots & y_{2,p} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{n,1} & \cdots & x_{n,q} & y_{n,1} & \cdots & y_{n,p} \end{pmatrix} \begin{array}{l} \leftarrow \text{data for } e_1 \\ \leftarrow \text{data for } e_2 \\ \vdots \\ \leftarrow \text{data for } e_n \end{array}$$
$$\begin{array}{cccccc} \uparrow & & \uparrow & \uparrow & & \uparrow \\ X_1 & \cdots & X_q & Y_1 & \cdots & Y_p \end{array}$$

where x_{i1}, \dots, x_{iq} and y_{i1}, \dots, y_{ip} are the data of X_1, \dots, X_q and Y_1, \dots, Y_p , respectively, for object e_i .

Estimating the Expectation Values: $\mu_{X_j} = E(X_j)$ and $\mu_{Y_j} = E(Y_j)$ are estimated by the **column means**

$$\mu_{X_j} = E(X_j) \approx \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad \text{and} \quad \mu_{Y_j} = E(Y_j) \approx \bar{y}_j = \frac{1}{n} \sum_{i=1}^n y_{ij}.$$

Data Matrix and Covariance Matrix of the Centered Data

Data Matrix of the Centered Data:

$$\mathbf{W} = (\mathbf{W}_x, \mathbf{W}_y)$$
$$= \begin{pmatrix} x_{1,1} - \bar{x}_1 & \cdots & x_{1,q} - \bar{x}_q & y_{1,1} - \bar{y}_1 & \cdots & y_{1,p} - \bar{y}_p \\ x_{2,1} - \bar{x}_1 & \cdots & x_{2,q} - \bar{x}_q & y_{2,1} - \bar{y}_1 & \cdots & y_{2,p} - \bar{y}_p \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{n,1} - \bar{x}_1 & \cdots & x_{n,q} - \bar{x}_q & y_{n,1} - \bar{y}_1 & \cdots & y_{n,p} - \bar{y}_p \end{pmatrix} \begin{matrix} \leftarrow e_1 \\ \leftarrow e_2 \\ \vdots \\ \leftarrow e_n \end{matrix}$$
$$\begin{matrix} \uparrow & & \uparrow & \uparrow & & \uparrow \\ X_1 - \mu_{X_1} & \cdots & X_q - \mu_{X_q} & Y_1 - \mu_{Y_1} & \cdots & Y_p - \mu_{Y_p} \end{matrix}$$

Covariance Matrix of the Data of the Measurement Variables:

$$\mathbf{S} = \frac{1}{n-1} \mathbf{W}' \mathbf{W} = \frac{1}{n-1} \begin{pmatrix} \mathbf{W}'_x \\ \mathbf{W}'_y \end{pmatrix} (\mathbf{W}_x, \mathbf{W}_y)$$
$$= \begin{pmatrix} \frac{1}{n-1} \mathbf{W}'_x \mathbf{W}_x & \frac{1}{n-1} \mathbf{W}'_x \mathbf{W}_y \\ \frac{1}{n-1} \mathbf{W}'_y \mathbf{W}_x & \frac{1}{n-1} \mathbf{W}'_y \mathbf{W}_y \end{pmatrix}$$

More on the Covariance Matrix of the Centered Data I

$\frac{1}{n-1} \mathbf{W}'_x \mathbf{W}_x$ is the covariance matrix of the data of $X_j, j = 1, \dots, q$.

$$\left(\frac{1}{n-1} \mathbf{W}'_x \mathbf{W}_x \right)_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j) (x_{ik} - \bar{x}_k) = \widehat{\text{Cov}}(X_j, X_k),$$

$$\left(\frac{1}{n-1} \mathbf{W}'_x \mathbf{W}_x \right)_{jj} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = \widehat{\text{Var}}(X_j)$$

$\mathbf{W}'_y \mathbf{W}_y$ is the covariance matrix of the data of $Y_j, j = 1, \dots, p$.

$$\left(\frac{1}{n-1} \mathbf{W}'_y \mathbf{W}_y \right)_{jk} = \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j) (y_{ik} - \bar{y}_k) = \widehat{\text{Cov}}(Y_j, Y_k),$$

$$\left(\frac{1}{n-1} \mathbf{W}'_y \mathbf{W}_y \right)_{jj} = \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2 = \widehat{\text{Var}}(Y_j)$$

Notation: The $\hat{\cdot}$ in $\widehat{\text{Cov}}$ and $\widehat{\text{Var}}$ indicates that the quantity is an estimator of the covariance Cov and variance Var, respectively.

More on the Covariance Matrix of the Centered Data II

$\mathbf{W}'_x \mathbf{W}_y = (\mathbf{W}'_y \mathbf{W}_x)'$ is the matrix with all covariances between a variable X_j , $j = 1, \dots, q$, and a variable Y_k , $k = 1, \dots, p$.

$$\left(\frac{1}{n-1} \mathbf{W}'_x \mathbf{W}_y \right)_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j) (y_{ik} - \bar{y}_k) = \widehat{\text{Cov}}(X_j, Y_k)$$

Ex. 4.2 (Empirical Covariance Matrix): Given the data below for the measurement variables $X_1 = \text{yearly salary in 1000 Euros}$, $Y_1 = \text{average hours of work per week}$, $Y_2 = \text{average number of lines of code per week (measured in units of 100 lines of code)}$, for a software programmer on a non-pay-scale salary, compute the **empirical covariance matrix S**.

Programmer	X_1	Y_1	Y_2
e_1	50	45	50
e_2	60	55	55
e_3	70	50	60

Solving the LISREL Model

Strategy: We now want to compare the empirical covariance matrix \mathbf{S} with the model covariance matrix Σ that is obtained from our LISREL model equations. (How to compute Σ is explained on the next slides.)

The model covariance matrix Σ will depend on the unknown parameters of our LISREL model:

Param. (structural model): γ_{ij} , β_{ij} , $\phi_{ij} = \text{Cov}(\xi_i, \xi_j)$, $\psi_{ij} = \text{Cov}(\zeta_i, \zeta_j)$

Param. (measurement model): λ_{ij}^X , λ_{ij}^Y , $\theta_{ij}^\delta = \text{Cov}(\delta_i, \delta_j)$, $\theta_{ij}^\varepsilon = \text{Cov}(\varepsilon_i, \varepsilon_j)$

- 1 Equating $\Sigma = \mathbf{S}$, under suitable circumstances we can identify all the model parameters.
- 2 Finally the values of the latent variables can be computed from the measurement models in the same way used to compute the values of the factors in factor analysis (i.e. via solving an OLS problem).

How to Compute the Model Covariance Matrix

We know (see 'Topic 1: Revision of Linear Algebra and Variance/Covariance') that for two random variables X and Y

$$\text{Cov}(X, Y) = E((X - \mu_X) \cdot (Y - \mu_Y)). \quad (15)$$

Applying (15) and the linearity of the expectation value (see (6) on page 32) to the covariances $\text{Cov}(X_i, X_j)$, $\text{Cov}(X_i, Y_j)$, $\text{Cov}(Y_i, Y_j)$ and using the reflective measurement equations

$$Y_i - \mu_{Y_i} = \lambda_{i1}^Y \eta_1 + \lambda_{i2}^Y \eta_2 + \dots + \lambda_{ir}^Y \eta_r + \varepsilon_i = \sum_{j=1}^r \lambda_{ij}^Y \eta_j + \varepsilon_i \quad (16)$$

$$X_i - \mu_{X_i} = \lambda_{i1}^X \xi_1 + \lambda_{i2}^X \xi_2 + \dots + \lambda_{is}^X \xi_s + \delta_i = \sum_{j=1}^s \lambda_{ij}^X \xi_j + \delta_i \quad (17)$$

and the model assumptions we obtain equations for these covariances in terms of the model parameters (example on next slide).

Computing the Entries of the Model Covariance Matrix I

$$\begin{aligned}
 \text{Cov}(X_i, X_j) &= E((X_i - \mu_{X_i}) \cdot (X_j - \mu_{X_j})) \\
 &= E\left(\left[\lambda_{i1}^X \xi_1 + \dots + \lambda_{is}^X \xi_s + \delta_i\right] \cdot \left[\lambda_{j1}^X \xi_1 + \dots + \lambda_{js}^X \xi_s + \delta_j\right]\right) \\
 &= E\left(\lambda_{i1}^X \lambda_{j1}^X \xi_1 \xi_1 + \dots + \lambda_{is}^X \lambda_{j1}^X \xi_s \xi_1 + \dots + \lambda_{i1}^X \lambda_{js}^X \xi_1 \xi_s \right. \\
 &\quad \left. + \dots + \lambda_{is}^X \lambda_{js}^X \xi_s \xi_s + \lambda_{i1}^X \xi_1 \delta_j + \dots + \lambda_{is}^X \xi_s \delta_j \right. \\
 &\quad \left. + \lambda_{j1}^X \delta_i \xi_1 + \dots + \lambda_{js}^X \delta_i \xi_s + \delta_i \delta_j\right) \\
 &= \lambda_{i1}^X \lambda_{j1}^X \underbrace{E(\xi_1 \xi_1)}_{= \text{Var}(\xi_1) = \phi_{1,1}} + \dots + \lambda_{is}^X \lambda_{j1}^X \underbrace{E(\xi_s \xi_1)}_{= \text{Cov}(\xi_s, \xi_1) = \phi_{s,1}} + \dots + \lambda_{i1}^X \lambda_{js}^X \underbrace{E(\xi_1 \xi_s)}_{= \text{Cov}(\xi_1, \xi_s) = \phi_{1,s}} \\
 &\quad + \dots + \lambda_{is}^X \lambda_{js}^X \underbrace{E(\xi_s \xi_s)}_{= \text{Var}(\xi_s) = \phi_{s,s}} + \lambda_{i1}^X \underbrace{E(\xi_1 \delta_j)}_{= \text{Cov}(\xi_1, \delta_j) = 0} + \dots + \lambda_{is}^X \underbrace{E(\xi_s \delta_j)}_{= \text{Cov}(\xi_s, \delta_j) = 0} \\
 &\quad + \lambda_{j1}^X \underbrace{E(\delta_i \xi_1)}_{= \text{Cov}(\delta_i, \xi_1) = 0} + \dots + \lambda_{js}^X \underbrace{E(\delta_i \xi_s)}_{= \text{Cov}(\delta_i, \xi_s) = 0} + \underbrace{E(\delta_i \delta_j)}_{= \text{Cov}(\delta_i, \delta_j) = \theta_{ij}^\delta}
 \end{aligned}$$

Computing the Entries of the Model Covariance Matrix II

and we find:

$$\begin{aligned} \text{Cov}(X_i, X_j) = & \lambda_{i1}^X \lambda_{j1}^X \phi_{1,1} + \dots + \lambda_{is}^X \lambda_{j1}^X \phi_{s,1} + \dots \\ & + \lambda_{i1}^X \lambda_{js}^X \phi_{1,s} + \dots + \lambda_{is}^X \lambda_{js}^X \phi_{s,s} + \theta_{ij}^\delta \end{aligned} \quad (18)$$

Next we would compute the other covariances $\text{Cov}(Y_i, Y_j)$ and $\text{Cov}(X_i, Y_j) = \text{Cov}(Y_j, X_i)$ and then equate them with the respective empirical covariances:

$$\text{Cov}(X_i, X_j) = \widehat{\text{Cov}}(X_i, X_j), \quad i, j = 1, \dots, q, \quad (19)$$

$$\text{Cov}(Y_i, Y_j) = \widehat{\text{Cov}}(Y_i, Y_j), \quad i, j = 1, \dots, p, \quad (20)$$

$$\text{Cov}(X_i, Y_j) = \widehat{\text{Cov}}(X_i, Y_j), \quad i = 1, \dots, q, j = 1, \dots, p. \quad (21)$$

This looks complicated, but for each i usually **only a few (often one) of the coefficients** $\lambda_{ij}^Y, j = 1, \dots, r$, in (16) and $\lambda_{ij}^X, j = 1, \dots, s$, in (17) are $\neq 0$.

We will **test this for a very simple toy example** of a SEM on slide 112.

Inspecting (18), we see that (although we have a linear LISREL model) the **covariance equations in the model parameters are in general nonlinear**.

Identifiable Model: A LISREL model is **identifiable**, if there exists a **unique and valid set of model parameters** such that the equations (19), (20) and (21) from equating the empirical and the model covariance matrix are true.

A **necessary (but not sufficient) condition** for the identifiability of a LISREL model is that the **number of equations equals or is larger than the number of the model parameters**. The number of covariance equations is

$$\frac{(p+q)^2 - (p+q)}{2} + (p+q) = \frac{(p+q+1)(p+q)}{2}.$$

- There are more complicated criteria that allow to determine whether a LISREL model is identifiable.
- What do we mean by a **valid** set of parameters? There are some **solutions for the model parameters that do not make sense**, such as solutions with negative variances. Such solutions are **not valid**.

Choosing the Scale of the Latent Variables

The latent variables do not come with a 'natural' scale, and therefore we can prescribe the scale in the following ways:

Choosing the Scale of Exogenous Latent Variables:

For the exogenous latent variables ξ_i , we prescribe $\phi_{ii} = \text{Var}(\xi) = 1$.

Alternatively, for each ξ_i we can choose one non-zero coefficient λ_{ji}^X to have the value one $\lambda_{ji}^X = 1$, i.e. the exogenous latent variable ξ_i has the same scale as one of its centered measurement variables $X_j - \mu_{X_j}$.*

Choosing the Scale of Endogenous Latent Variables:

For the endogenous latent variables η_i , we choose one non-zero coefficient λ_{ji}^Y to have the value one $\lambda_{ji}^Y = 1$, i.e. the endogenous latent variable η_i has the same scale as one of its measurement variables $Y_j - \mu_{Y_j}$.*

* Choose a coefficient for having value one that you expect to be positive.

Restrictions on Model Parameters

Parameters in the LISREL model can be restricted if this is supported by logical cause-effect considerations:

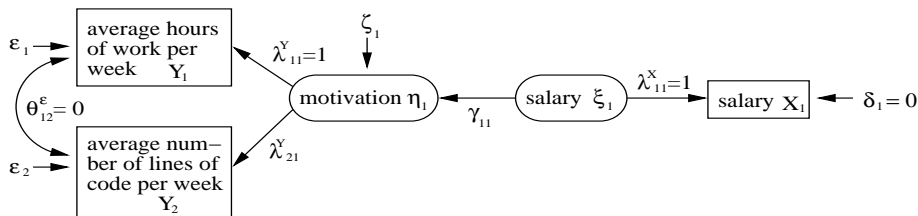
- 1 Individual error terms δ_i or ε_j can be set to zero if X_i or Y_j , respectively, are measured exactly.
- 2 Individual coefficients can be set to be equal, e.g. $\lambda_{1,1}^X = \lambda_{2,1}^X = a$, where a is some number that still needs to be determined.
- 3 Individual coefficients λ_{ij}^X or λ_{ij}^Y can be set to equal a fixed value.
- 4 Individual covariances ϕ_{ij} , ψ_{ij} , θ_{ij}^δ or θ_{ij}^ε can be set to be zero.

Attention:

- Any restrictions of parameters need to be founded on logical cause-effect considerations!
- Each type of restriction reduces the number of (free) parameters by one. Consequently the restriction of some parameters can make a LISREL model identifiable.

Ex. 4.3: Solving a Structural Equation Model with LISREL

To demonstrate the solution of a structural equation model with LISREL, we consider the simplified model for the **work of a software programmer on a non-pay-scale salary** shown in the diagram below.



- 1 Set up the structural equation model by specifying the **structural (inner) model** and the **(outer) measurement model**.
- 2 Determine with the LISREL approach **the model parameters** in terms of the covariances of the measurement variables.
- 3 Use the empirical covariance matrix from Ex. 4.2 to compute the **numerical values for the parameters** and interpret your results.

How the LISREL Model is Computed in Practice

For complicated LISREL Models it is **not** feasible to equate the empirical and model covariance matrix and solve for the parameters.

Minimization Problem: Let τ denote the vector of all model parameters of the LISREL model. A **discrepancy function** $F(\mathbf{S}, \mathbf{\Sigma}(\tau))$ for measuring the distance between the model covariance matrix $\mathbf{\Sigma}(\tau)$ and the empirical covariance matrix \mathbf{S} is **minimized**:

$$\text{Find } \tau_0 \text{ such that } F(\mathbf{S}, \mathbf{\Sigma}(\tau_0)) = \min_{\tau} F(\mathbf{S}, \mathbf{\Sigma}(\tau)),$$

where the **minimization** is over all valid choices of the parameter vector τ .

Ordinary Least Squares Discrepancy Function:

$$F(\mathbf{S}, \mathbf{\Sigma}(\tau)) = \frac{1}{2} \sum_{i=1}^{p+q} \sum_{j=1}^{p+q} [\mathbf{S}_{ij} - (\mathbf{\Sigma}(\tau))_{ij}]^2 = \frac{1}{2} \left(\begin{array}{l} \text{sum of squares of all} \\ \text{entries of } \mathbf{S} - \mathbf{\Sigma}(\tau) \end{array} \right)$$

Generalized Least Squares Discrepancy Function:

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\tau})) = \frac{1}{2} \sum_{i=1}^{p+q} \sum_{j=1}^{p+q} \left[([\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\tau})] \mathbf{S}^{-1})_{ij} \right]^2 = \frac{1}{2} \left(\begin{array}{l} \text{sum of squares} \\ \text{of all entries of} \\ [\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\tau})] \mathbf{S}^{-1} \end{array} \right)$$

Maximum Likelihood (ML) Discrepancy Function:

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\tau})) = \ln(\det(\boldsymbol{\Sigma}(\boldsymbol{\tau}))) + \text{tr}(\mathbf{S}(\boldsymbol{\Sigma}(\boldsymbol{\tau}))^{-1}) - \ln(\det(\mathbf{S})) - (p + q)$$

How is the minimization problem solved?

An iterative algorithm is used that computes a sequence of approximations $\boldsymbol{\tau}_k$, $k = 1, 2, \dots$, of the parameter vector $\boldsymbol{\tau}_0$, such that

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\tau}_{k+1})) \leq F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\tau}_k)).$$

The iterative algorithm stops when $\text{dist}(\boldsymbol{\tau}_{k+1}, \boldsymbol{\tau}_k)$ is very small, i.e. when the parameter vectors change only very little in subsequent iterative steps.

Evaluation of a LISREL Model: Quality Criteria

There is a variety of **criteria for evaluating how well a LISREL model** (as a whole or also its individual parts) **models the data**.

This goes beyond the scope of this course, but if you work with structural equation models, you will have to learn about these quality criteria.

Some words of general advice and a first assessment of your model:

- **Do the path coefficients** in the inner model and in the measurement models **have the expected signs**? If not your LISREL model supports causal relationships that are **contrary** to your expected cause-effect relationships!
- **Do you get negative variances or other effects that cannot occur in a correct model**? If so, this may indicate a badly specified model or numerical problems in solving the LISREL model.

LISREL: Structural (Inner) Model in Matrix Form

$\xi = (\xi_1, \dots, \xi_s)'$ = vector of exogenous (independent) latent variables ξ_i

$\eta = (\eta_1, \dots, \eta_r)'$ = vector of endogenous (dependent) latent variables η_i

$\zeta = (\zeta_1, \dots, \zeta_r)'$ = vector of error variables ζ_i in regression equations for η_i

$\Gamma = (\gamma_{ij})$ = $r \times s$ matrix with the coefficients γ_{ij} for the effect of ξ_j on η_i

$B = (\beta_{ij})$ = $r \times r$ matrix with the coefficients β_{ij} for the effect of η_j on η_i ,
where $\beta_{ii} = 0$ for all $i = 1, 2, \dots, r$

Structural (Inner) Model in Matrix Form: $\eta = B\eta + \Gamma\xi + \zeta$

Assumptions: $E(\xi) = \mathbf{0}$, $E(\eta) = \mathbf{0}$, $E(\zeta) = \mathbf{0}$, $\text{Cov}(\xi, \zeta) = \mathbf{0}$.

Parameters: coefficient matrices $B = (\beta_{ij})$, $\Gamma = (\gamma_{ij})$ and covariance matrices $\Psi = (\text{Cov}(\zeta_i, \zeta_j)) = \text{Cov}(\zeta, \zeta)$, $\Phi = (\text{Cov}(\xi_i, \xi_j)) = \text{Cov}(\xi, \xi)$

Other Unknowns: values of latent variables, $\text{Cov}(\eta_i, \eta_j)$, $i, j = 1, \dots, r$

$\mathbf{y} = (Y_1, \dots, Y_p)'$ = vector of measurement variables for η_1, \dots, η_r

$\boldsymbol{\mu}_y = (\mu_{Y_1}, \dots, \mu_{Y_p})'$ = vector of expectation values of Y_1, \dots, Y_p

$\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p)'$ = vector of measurement errors for Y_1, \dots, Y_p

$\boldsymbol{\Lambda}_Y = (\lambda_{ij}^Y) = p \times r$ matrix with coefficients λ_{ij}^Y for the effect of η_j on Y_i

$\mathbf{x} = (X_1, \dots, X_q)'$ = measurement variables for ξ_1, \dots, ξ_s

$\boldsymbol{\mu}_x = (\mu_{X_1}, \dots, \mu_{X_q})'$ = vector of expectation values of X_1, \dots, X_q

$\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)'$ = vector of measurement errors for X_1, \dots, X_q

$\boldsymbol{\Lambda}_X = (\lambda_{ij}^X) = q \times s$ matrix with coefficients λ_{ij}^X for the effect of ξ_j on X_i

Factor Analytical Equations:

$$\mathbf{y} - \boldsymbol{\mu}_y = \boldsymbol{\Lambda}_Y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (22)$$

$$\mathbf{x} - \boldsymbol{\mu}_x = \boldsymbol{\Lambda}_X \boldsymbol{\xi} + \boldsymbol{\delta} \quad (23)$$

Assumptions:

- $E(\varepsilon) = \mathbf{0}$ and $E(\delta) = \mathbf{0}$
- $\text{Cov}(\varepsilon, \eta) = \mathbf{0}$, $\text{Cov}(\varepsilon, \xi) = \mathbf{0}$, $\text{Cov}(\delta, \eta) = \mathbf{0}$ and $\text{Cov}(\delta, \xi) = \mathbf{0}$
- $\text{Cov}(\varepsilon, \delta) = \mathbf{0}$
- $\text{Cov}(\varepsilon, \zeta) = \mathbf{0}$ and $\text{Cov}(\delta, \zeta) = \mathbf{0}$

Parameters: coefficient matrices Λ_X , Λ_Y and the covariance matrices $\Theta_\varepsilon = (\theta_{ij}^\varepsilon) = \text{Cov}(\varepsilon, \varepsilon)$ with $\theta_{ij}^\varepsilon = \text{Cov}(\varepsilon_i, \varepsilon_j)$, and $\Theta_\delta = (\theta_{ij}^\delta) = \text{Cov}(\delta, \delta)$ with $\theta_{ij}^\delta = \text{Cov}(\delta_i, \delta_j)$

Model Covariance Matrix: In matrix notation our model covariance matrix can be decomposed into four blocks:

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \quad \text{where} \quad \Sigma_{yx} = \Sigma'_{xy}.$$

Model Covariance Matrix

$$\begin{aligned}\Sigma_{xx} &= \text{Cov}(\mathbf{x}, \mathbf{x}) = E((\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)') \\ \Sigma_{xy} = \Sigma'_{yx} &= \text{Cov}(\mathbf{x}, \mathbf{y}) = E((\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{y} - \boldsymbol{\mu}_y)') \\ \Sigma_{yy} &= \text{Cov}(\mathbf{y}, \mathbf{y}) = E((\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{y} - \boldsymbol{\mu}_y)')\end{aligned}\quad (24)$$

Using (22) and (23) we can now compute (24) in terms of the model parameters. E.g. using (23) we find

$$\begin{aligned}\Sigma_{xx} &= E((\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)') \\ &= E([\boldsymbol{\Lambda}_X \boldsymbol{\xi} + \boldsymbol{\delta}][\boldsymbol{\Lambda}_X \boldsymbol{\xi} + \boldsymbol{\delta}]') = E([\boldsymbol{\Lambda}_X \boldsymbol{\xi} + \boldsymbol{\delta}][\boldsymbol{\xi}' \boldsymbol{\Lambda}'_X + \boldsymbol{\delta}']) \\ &= E(\boldsymbol{\Lambda}_X \boldsymbol{\xi} \boldsymbol{\xi}' \boldsymbol{\Lambda}'_X + \boldsymbol{\delta} \boldsymbol{\xi}' \boldsymbol{\Lambda}'_X + \boldsymbol{\Lambda}_X \boldsymbol{\xi} \boldsymbol{\delta}' + \boldsymbol{\delta} \boldsymbol{\delta}') \\ &= \boldsymbol{\Lambda}_X \underbrace{E(\boldsymbol{\xi} \boldsymbol{\xi}')}_{= \text{Cov}(\boldsymbol{\xi}, \boldsymbol{\xi}) = \boldsymbol{\Phi}} \boldsymbol{\Lambda}'_X + \underbrace{E(\boldsymbol{\delta} \boldsymbol{\xi}')}_{= \text{Cov}(\boldsymbol{\delta}, \boldsymbol{\xi}) = 0} \boldsymbol{\Lambda}'_X + \boldsymbol{\Lambda}_X \underbrace{E(\boldsymbol{\xi} \boldsymbol{\delta}')}_{= \text{Cov}(\boldsymbol{\xi}, \boldsymbol{\delta}) = 0} + \underbrace{E(\boldsymbol{\delta} \boldsymbol{\delta}')}_{= \text{Cov}(\boldsymbol{\delta}, \boldsymbol{\delta}) = \boldsymbol{\Theta}_\delta} \\ &= \boldsymbol{\Lambda}_X \boldsymbol{\Phi} \boldsymbol{\Lambda}'_X + \boldsymbol{\Theta}_\delta\end{aligned}$$

where we have used the fact that the expectation value is linear to pull the matrix $\boldsymbol{\Lambda}_X$ out the expectation values and have also used the assumptions on the previous pages.

Structural Equation Modeling

Topic 5: PLS Path Modeling (Partial Least Squares Path Modeling)

Dr. Kerstin Hesse

Email: kerstin.hesse@hhl.de; *Phone:* +49 (0)341 9851-820; *Office:* HHL Main Building, Room 115A

HHL Leipzig Graduate School of Management, Jahnallee 59, 04109 Leipzig, Germany

Doctoral Program at HHL, June 1-2, 2012

Topic 5: PLS (Partial Least Squares) Path Modeling

- Introduction to the PLS Path Model
 - some facts about PLS
 - mathematical properties of PLS
 - setting up the structural model for PLS path modeling
- Numerical Computation of the PLS Path Model: The PLS Algorithm
 - data matrices for the measurement blocks
 - An iteration step of the iterative algorithm has 3 substeps:
 - Step 1: outer estimation stage – computing values for the latent variables as a weighted sum of the measurement variables
 - Step 2: inner estimation stage – computing new values for the latent variables as weighted sum of related latent variables
 - Step 3: computing new weights (for the weighted sum in step 1) – solving an ordinary least squares problem for each measurement block
 - After the iterative algorithm terminates we compute the final values of the latent variables and the path coefficients in the inner model.

Some Facts & Recommendations from the Literature

PLS: PLS Path Modeling is often called just PLS.

Why is it Called 'Partial Least Squares'? In each iterative step of PLS we solve a sequence of **local (partial) ordinary least squares problems**.

- PLS can be used to model causal models with either **formative or reflective** measurement models/blocks or a **mixture of both**.
- PLS makes **no distributional assumptions** but gives also less information about the model quality (than LISREL).
- PLS aims to **maximize the variance** of the variables **explained** by the model.
- PLS is considered more useful for **exploratory** purposes and less for confirmatory purposes.
- Because PLS makes no distributional assumptions, it is called a **soft modeling approach**.

Mathematical Features of PLS Path Modeling

Iterative Algorithm: PLS is an **iterative algorithm**, i.e. we **execute certain steps repeatedly** until some convergence criterion is satisfied (or until we have reached the maximum number of iteration steps).

Does the PLS Algorithm Converge? There is no **formal proof of convergence** for more than two measurement blocks, but **in most practical applications PLS appears to converge**.

Example Iterative Algorithm: Set $x = 0$.

For $n = 1, 2, \dots$ do the following

$$\left. \begin{array}{l} \text{set } x^{\text{old}} = x \\ \text{compute } x = x^{\text{old}} + \frac{1}{2^n} \end{array} \right\} \text{iteration step}$$

until $|x - x^{\text{old}}| < 0.01$ (stopping criterion)

This algorithm converges to the number 2. If we replace $x = x^{\text{old}} + \frac{1}{2^n}$ by $x = x^{\text{old}} + 1$, then the algorithm does not converge (it diverges).

Latent Variables and Structural (Inner) Model in PLS

- The **latent variables** are ξ_q , $q = 1, \dots, Q$. We do **not** distinguish in the notation between exogenous and endogenous latent variables.
 - Latent variables are **centered**, i.e. $E(\xi_q) = 0$ for all $q = 1, \dots, Q$.
-

The **regression equations** in the structural (inner) model can be written as

$$\xi_j = \sum_{\substack{q=1, \\ q:\xi_q \rightarrow \xi_j}}^Q \beta_{jq} \xi_q + \zeta_j, \quad j = 1, \dots, Q, \quad (25)$$

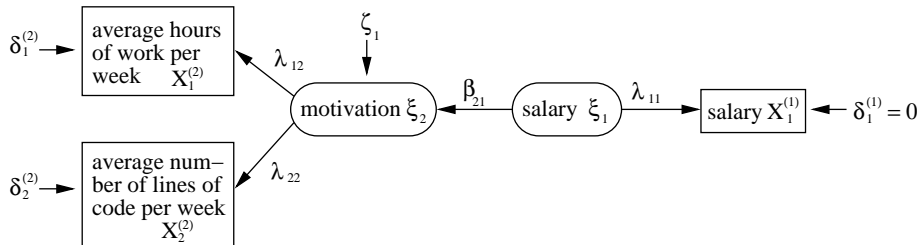
where $q : \xi_q \rightarrow \xi_j$ means that we **only sum over those indices q for which we have an arrow pointing from ξ_q to ξ_j in our path diagram.**

As usual ζ_j is the **error term** for ξ_j , satisfying $E(\zeta_j) = 0$. β_{jq} is the **coefficient** for the effect of ξ_q on ξ_j ; it is assumed that $\beta_{jq} \neq 0$.

Note: (25) is the **same equation** as in LISREL; we only **exclude in notation directly any zero coefficients**.

Example: Structural (Inner) Model with PLS

Example: SEM diagram for PLS for the work of a software programmer on a non-pay-scale salary.



Regression equations in PLS for the structural (inner) model:

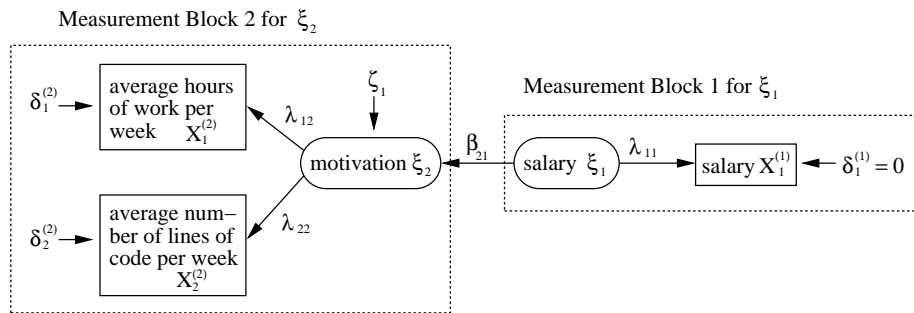
$$\xi_2 = \beta_{2,1} \xi_1 + \zeta_1 \quad (26)$$

Note: Inspecting the SEM diagram above, we see that the notation has also changed: **Measurement variables are now organized in blocks.**

Measurement Blocks in PLS

Measurement Block of a Latent Variable: Each latent variable ξ_q has a **measurement block** consisting of all its P_q measurement variables which we name $X_1^{(q)}, \dots, X_{P_q}^{(q)}$ (upper index (q) for the latent variable ξ_q).

Example (continued): We have indicated the measurement blocks in the SEM diagram of our example.



Example: Reflective Measurement Blocks in PLS

Measurement Block for ξ_1 : The factor analytic equation is

$$X_1^{(1)} - \mu_{X_1^{(1)}} = \lambda_{1,1} \xi_1 + \delta_1^{(1)}, \quad (27)$$

where: $\mu_{X_1^{(1)}}$ is the expectation value of $X_1^{(1)}$; $\delta_1^{(1)}$ is error term;
 $\lambda_{1,1}$ is the coefficient for the effect of ξ_1 on $X_1^{(1)}$.

Measurement Block for ξ_2 : The two factor analytic equations are

$$X_1^{(2)} - \mu_{X_1^{(2)}} = \lambda_{1,2} \xi_2 + \delta_1^{(2)}, \quad (28)$$

$$X_2^{(2)} - \mu_{X_2^{(2)}} = \lambda_{2,2} \xi_2 + \delta_2^{(2)}, \quad (29)$$

where: $\mu_{X_1^{(2)}}$ and $\mu_{X_2^{(2)}}$ are the expectation values of $X_1^{(2)}$ and $X_2^{(2)}$;
 $\delta_1^{(2)}$ and $\delta_2^{(2)}$ are the error terms; $\lambda_{1,2}$ is the coefficient for the effect of ξ_2 on $X_1^{(2)}$; $\lambda_{2,2}$ is the coefficient for the effect of ξ_2 on $X_2^{(2)}$.

Measurement Variable Notation (for Both Measurement Models):

All measurement variables are written as $X_p^{(q)}$, where the upper index (q) indicates that $X_p^{(q)}$ measures the latent variable ξ_q . For each latent variable ξ_q , we number its P_q measurement variables:

$$X_p^{(q)}, \quad p = 1, 2, \dots, P_q.$$

Formative Measurement Block for ξ_q :

$$\xi_q = \sum_{p=1}^{P_q} \lambda_{qp} \left(X_p^{(q)} - \mu_{X_p^{(q)}} \right) + \delta_q,$$

where $\mu_{X_p^{(q)}}$ is the expectation value of $X_p^{(q)}$, and where δ_q is error term.

The coefficient λ_{qp} is the coefficient for the effect of $X_p^{(q)}$ on ξ_q .

Assumptions: $E(\delta_q) = 0$ and $\text{Cov}(\delta_q, X_p^{(q)}) = 0$ for all $p = 1, \dots, P_q$.

Reflective Measurement Block for ξ_q :

$$X_p^{(q)} - \mu_{X_p^{(q)}} = \lambda_{pq} \xi_q + \delta_p^{(q)}, \quad p = 1, 2, \dots, P_q,$$

where $\mu_{X_p^{(q)}}$ is the expectation value of $X_p^{(q)}$, and where $\delta_p^{(q)}$ is error term.

The coefficient λ_{pq} is the coefficient for the effect of ξ_q on $X_p^{(q)}$.

Assumptions: $E(\delta_p^{(q)}) = 0$ and $\text{Cov}(\delta_p^{(q)}, \xi_q) = 0$ for all $p = 1, \dots, P_q$.

Requirements on a Reflective Measurement Block for ξ_q :

- **unidimensionality:** In a factor analytic model for the measurement variables in the block, all variables in that block are mainly explained the factor ξ_q .
- **homogeneity:** The correlations between ξ_q and its measurement variables $X_p^{(q)}$ have all the same sign.

MIMIC Model: A MIMIC model is a structural equation model with some reflective and some formative measurement blocks. PLS can handle MIMIC models.

Sampling Data:

- All measurement variables are observed for the same N objects e_1, \dots, e_N (regardless to which measurement block these measurement variables belong).
- Data is organized by measurement block: in the block for ξ_q we have

$$x_{np}^{(q)} = \text{value of random variable } X_p^{(q)} \text{ for object } e_n$$

where $n = 1, 2, \dots, N$ (for the N objects) and $p = 1, 2, \dots, P_q$ (for the P_q measurement variables in the block for ξ_q)

PLS Algorithm: The Iterative Algorithm

One Iterative Step:

- 1 We compute the latent variables, by taking a weighted sum (linear combination) of the corresponding measurement variables.
- 2 We compute new latent variables as weighted sums of the related latent variables. (A latent variable is related to another latent variable, if it is connected with this latent variable by an arrow.)
- 3 We compute new weights for the sum in step 1: This is done for each measurement block separately by solving an ordinary least squares problem. There are two ways for computing the new weights that roughly correspond a formative or a reflective measurement block.

After step 3 we start again with step 1, now using the new weights (that were computed in step 3) in the new step 1.

Stopping criterion: Once the weights from step 3 change only very little from one iterative step to the next, we stop the iterative algorithm.

Final Latent Variables:

Once the iterative algorithm has stopped, we use the weights from the last step 3, and **execute step 1 once more to compute the final latent variables** from the measurement variables of the corresponding block.

Computation of the Path Coefficients:

With the final latent variables, we use **regression (i.e. ordinary least squares)** to **compute the path coefficients of the structural (inner) model**.

Remark on the Algorithm in General: For brevity we talk in the algorithm about the measurement variables and latent variables. The computations are of course done with **the data for these variables** for the objects e_1, e_2, \dots, e_N in the sample.

Iterative Algorithm in PLS: Input and Explanation

Input:

- for each $q = 1, \dots, Q$, the **centered data** for the q th measurement block for ξ_q : $x_{n,p}^{(q)} - \bar{x}_p^{(q)}$, $p = 1, \dots, P_q$, $n = 1, \dots, N$
 - **initial weight vectors** $\mathbf{w}^{(q)} = (w_1^{(q)}, \dots, w_{P_q}^{(q)})'$, $q = 1, \dots, Q$, for the Q measurement blocks, i.e. $\mathbf{w}^{(q)}$ = weight vector for block for ξ_q
(The **initial weights** could for example be chosen all equal, i.e. $\mathbf{w}^{(q)} = (1/P_q, \dots, 1/P_q)'$ with $w_p^{(q)} = 1/P_q$ for $p = 1, \dots, P_q$.)
-

Iterative Algorithm Explained on Two Levels:

We explain the algorithm in parallel as an **algorithm for the random variables** and as an **algorithm for the measured data from the sample**.

- The algorithm for the random variables is **more intuitive** but we cannot compute anything.
- The algorithm for the measured data **transforms** the algorithm for the random variables into a computable version for the measured data.

Iterative Algorithm – Step 1: Outer Estimation Stage

Random Variable Level:

For $q = 1, \dots, Q$ the latent variable ξ_q is **estimated** via the standardized variable $\hat{\xi}_q = (\eta_q - \mu_{\eta_q}) / \sigma_{\eta_q}$, where

$$\eta_q = \pm \sum_{p=1}^{P_q} w_p^{(q)} \left(X_p^{(q)} - \mu_{X_p^{(q)}} \right), \quad (30)$$

i.e. we approximate ξ_q by a standardized linear combination of its measurement variables $X_1^{(q)}, \dots, X_{P_q}^{(q)}$.

Data Level:

For each object e_n , $n = 1, \dots, N$, and each $q = 1, \dots, Q$, we compute (following (30)) a value for η_q

$$\eta_{nq} = \pm \sum_{p=1}^{P_q} w_p^{(q)} \left(x_{np}^{(q)} - \overline{x_p^{(q)}} \right). \quad (31)$$

Iterative Algorithm – Step 1: Outer Estimation Stage

Then we **standardize the data** for the random variable η_q , i.e.

$$\bar{\eta}_q = \frac{1}{N} \sum_{n=1}^N \eta_{nq} \quad \text{and} \quad s_{\eta_q} = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (\eta_{nq} - \bar{\eta}_q)^2}$$

and the data of the estimator $\hat{\xi}_q$ of ξ_q for object e_n is given by

$$\xi_{nq} = \frac{\eta_{nq} - \bar{\eta}_q}{s_{\eta_q}}.$$

Choice of the Sign in (30) and (31): The **sign** in (30) and (31) is chosen such that η_q is **positively correlated with the majority of its measurement variables** $X_1^{(q)}, X_2^{(q)}, \dots, X_{P_q}^{(q)}$ (i.e. at least 50% of the non-zero empirical covariances $\widehat{\text{Cov}}(\eta_q, X_p^{(q)})$, $p = 1, \dots, P_q$, are positive).

We note that η_q and its data satisfy by design $\mu_{\eta_q} = 0$ and $\bar{\eta}_q = 0$.

Ex. 5.1 (a) Step 1 of the Iterative Algorithm

For the SEM in the example on pages 125–127 (which describes the work of a software programmer on a non-pay-scale salary) with the model equations (26) to (29), we are given the following data

Programmer	$X_1^{(1)}$	$X_1^{(2)}$	$X_2^{(2)}$
e_1	50	45	50
e_2	60	55	55
e_3	70	50	60

for the measurement variables: $X_1^{(1)}$ = yearly salary in 1000 Euros, $X_1^{(2)}$ = average hours of work per week, $X_2^{(2)}$ = average number of lines of code per week (measured in units of 100 lines of code). **Using equal weights as the initial weights, execute step 1 of the PLS algorithm.**

Iterative Algorithm – Step 2: Inner Estimation Stage

Random Variable Level: For $q = 1, \dots, Q$, the latent variable ξ_q is approximated by the standardized variable $\nu_q = (\rho_q - \mu_{\rho_q})/\sigma_{\rho_q}$, where

$$\rho_q = \sum_{\substack{j=1, \\ q: \xi_q \rightarrow \xi_j, \\ \text{or } q: \xi_j \rightarrow \xi_q}}^Q e_{qj} \hat{\xi}_j. \quad (32)$$

The summation is over all latent variables that are connected with ξ_q via an arrow. We approximate ξ_q by a standardized linear combination of the estimators $\hat{\xi}_j$ (computed in step 1) of those random variables that are connected with ξ_q .

Schemes for Choosing the Weights e_{qj} (see also page 139):

- Centroid scheme: $e_{qj} = \text{sign of Cov}(\hat{\xi}_q, \hat{\xi}_j)$
- Factorial scheme: $e_{qj} = (\text{correlation between } \hat{\xi}_q \text{ and } \hat{\xi}_j) = \text{Cov}(\hat{\xi}_q, \hat{\xi}_j)$
- Path weighting scheme: see literature

Iterative Algorithm – Step 2: Inner Estimation Stage

Data Level: For each object e_n , $n = 1, 2, \dots, N$, and for $q = 1, \dots, Q$, we compute (analogously to (32)) a **value for** ρ_q :

$$\rho_{nq} = \sum_{\substack{j=1, \\ q: \xi_q \rightarrow \xi_j, \\ \text{or } q: \xi_j \rightarrow \xi_q}}^Q e_{qj} \xi_{nj}. \quad (33)$$

Then we **standardize the data for the random variable** ρ_q , i.e.

$$\overline{\rho_q} = \frac{1}{N} \sum_{n=1}^N \rho_{nq} \quad \text{and} \quad s_{\rho_q} = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (\rho_{nq} - \overline{\rho_q})^2}$$

and the data for object e_n for ν_q is given by

$$\nu_{nq} = \frac{\rho_{nq} - \overline{\rho_q}}{s_{\rho_q}}.$$

We note that by design $E(\rho_q) = 0$ and $\overline{\rho_q} = 0$.

Practical Computation of the Weights e_{qj} :

The weights must e_{qj} must be computed from the **empirical covariance** of $\hat{\xi}_q$ and $\hat{\xi}_j$: $e_{qj} = \text{sign of } \widehat{\text{Cov}}(\hat{\xi}_q, \hat{\xi}_j)$ (for the centroid scheme) or $e_{qj} = \widehat{\text{Cov}}(\hat{\xi}_q, \hat{\xi}_j)$ (for the factorial scheme), where

$$\widehat{\text{Cov}}(\hat{\xi}_q, \hat{\xi}_j) = \frac{1}{N-1} \sum_{n=1}^N (\xi_{nq} - \underbrace{\bar{\xi}_q}_{=0}) (\xi_{nj} - \underbrace{\bar{\xi}_j}_{=0}) = \frac{1}{N-1} \sum_{n=1}^N \xi_{nq} \xi_{nj}$$

which is just the **correlation (coefficient)** of $\hat{\xi}_q$ and $\hat{\xi}_j$ because $\hat{\xi}_q$ and $\hat{\xi}_j$ are standardized.

Ex. 5.1 (b) Step 2 of the Iterative Algorithm

Using the results from Ex. 5.1 (a) for the structural equation model given in Ex. 5.1 (a), **execute step 2 of the iterative algorithm** with the **centroid weights scheme**.

Iterative Algorithm – Step 3: New Weights, Mode A

Mode A for the computation of the new weights $w_p^{(q)}$ is more appropriate for PLS with **reflective measurement models**.

Mode A is also suggested for **endogenous** latent variables.

Mode A – Random Variable Level:

$$w_p^{(q)} = \text{Cov}(X_p^{(q)}, \nu_q) = \left(\begin{array}{l} \text{covariance of the measurement} \\ \text{variable } X_p^{(q)} \text{ and the approximation } \nu_q \\ \text{of the latent variable } \xi_q \text{ from step 2} \end{array} \right)$$

Mode A – Data Level: From the data for the objects e_1, e_2, \dots, e_N , we compute the new weights $w_p^{(q)}$ via

$$w_p^{(q)} = \widehat{\text{Cov}}(X_p^{(q)}, \nu_q) = \frac{1}{N-1} \sum_{n=1}^N \left(x_{np}^{(q)} - \overline{x_p^{(q)}} \right) \nu_{nq}. \quad (34)$$

Iterative Algorithm – Step 3: Explanation of Mode A

Consider the **reflective measurement equation** for measuring ξ_q with $X_p^{(q)}$:

$$X_p^{(q)} - \mu_{X_p^{(q)}} = \lambda_{pq} \nu_q + \delta_p^{(q)},$$

where we have replaced ξ_q by its approximation ν_q .

We have data for ν_q and $X_p^{(q)} - \mu_{X_p^{(q)}}$, and finding the coefficient λ_{pq} becomes a **regression (OLS) problem**: The **minimization problem**

$$\min_{\lambda_{pq}} \left(\sum_{n=1}^N \left[\left(x_{np}^{(q)} - \overline{x_p^{(q)}} \right) - \lambda_{pq} \nu_{nq} \right]^2 \right)$$

leads to the formula (use calculus)

$$\sum_{n=1}^N \left(x_{np}^{(q)} - \overline{x_p^{(q)}} \right) \nu_{nq} = \lambda_{pq} \underbrace{\sum_{n=1}^N (\nu_{nq})^2}_{=N-1},$$

and after division by $N - 1$ we find (34) with $w_p^{(q)} = \lambda_{pq}$.

Ex. 5.1 (c) Step 3 (Mode A) of the Iterative Algorithm

Using the results from Ex. 5.1 (a) to (b) for the structural equation model given in Ex. 5.1 (a), execute step 3 of the iterative algorithm.

Centered Data Matrix for the Measurement Block for ξ_q :

$$\begin{aligned}
 \mathbf{Y}^{(q)} &= \left(x_{np}^{(q)} - \overline{x_p^{(q)}} \right)_{\substack{n=1, \dots, N \\ p=1, \dots, P_q}} \\
 &= \begin{pmatrix} x_{1,1}^{(q)} - \overline{x_1^{(q)}} & x_{1,2}^{(q)} - \overline{x_2^{(q)}} & \dots & x_{1,P_q}^{(q)} - \overline{x_{P_q}^{(q)}} \\ x_{2,1}^{(q)} - \overline{x_1^{(q)}} & x_{2,2}^{(q)} - \overline{x_2^{(q)}} & \dots & x_{2,P_q}^{(q)} - \overline{x_{P_q}^{(q)}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1}^{(q)} - \overline{x_1^{(q)}} & x_{N,2}^{(q)} - \overline{x_2^{(q)}} & \dots & x_{N,P_q}^{(q)} - \overline{x_{P_q}^{(q)}} \end{pmatrix} \begin{matrix} \leftarrow e_1 \\ \leftarrow e_2 \\ \vdots \\ \leftarrow e_N \end{matrix} \\
 &\quad \begin{matrix} \uparrow & \uparrow & & \uparrow \\ X_1^{(q)} - \mu_{X_1^{(q)}} & X_2^{(q)} - \mu_{X_2^{(q)}} & \dots & X_{P_q}^{(q)} - \mu_{X_{P_q}^{(q)}} \end{matrix}
 \end{aligned}$$

Iterative Algorithm – Step 3: New Weights, Mode B

Mode B for the computation of the new weights is more appropriate for PLS with **formative measurement models**.

Mode B is also suggested for **exogenous** latent variables.

Mode B – Random Variable Level: This is only relevant for the derivation of the formula for the weights (see next page).

Mode B – Data Level:

With $\mathbf{w}^{(q)} = (w_1^{(q)}, \dots, w_{P_q}^{(q)})'$ and $\boldsymbol{\nu}_q = (\nu_{1q}, \dots, \nu_{nq})'$ compute the weights $w_p^{(q)}$ via the **ordinary least squares (OLS)** equation

$$\mathbf{w}^{(q)} = [(\mathbf{Y}^{(q)})'(\mathbf{Y}^{(q)})]^{-1} (\mathbf{Y}^{(q)})' \boldsymbol{\nu}_q. \quad (35)$$

Iterative Algorithm – Step 3: Explanation of Mode B

Consider the **formative measurement equation** for measuring the measurement variables ξ_q with $X_1^{(q)}, \dots, X_{P_q}^{(q)}$:

$$\nu_q = \sum_{p=1}^{P_q} \lambda_{qp} (X_p^{(q)} - \mu_{X_p^{(q)}}) + \delta_q,$$

where we have replaced ξ_q by its approximation ν_q .

We have data for ν_q and $X_p^{(q)} - \mu_{X_p^{(q)}}$, and finding the coefficient λ_{pq} becomes a **regression (OLS) problem**: The **minimization problem**

$$\min_{\lambda_{pq}, \text{ where } p=1, \dots, P_q} \left(\sum_{n=1}^N \left[\nu_{nq} - \sum_{p=1}^{P_q} \lambda_{qp} (x_{np}^{(q)} - \overline{x_p^{(q)}}) \right]^2 \right)$$

leads (use calculus) to the linear system $(\mathbf{Y}^{(q)})'(\mathbf{Y}^{(q)}) \boldsymbol{\lambda}^{(q)} = (\mathbf{Y}^{(q)})' \boldsymbol{\nu}_q \Leftrightarrow \boldsymbol{\lambda}^{(q)} = [(\mathbf{Y}^{(q)})'(\mathbf{Y}^{(q)})]^{-1} (\mathbf{Y}^{(q)})' \boldsymbol{\nu}_q$, where $\boldsymbol{\lambda}^{(q)} = (\lambda_{q1}, \dots, \lambda_{qP_q})'$ and $\boldsymbol{\nu}_q = (\nu_{1q}, \dots, \nu_{nq})'$. We get (35) with $\mathbf{w}^{(q)} = \boldsymbol{\lambda}^{(q)}$.

Stopping Criterion for the Iterative Algorithm

Test Quantity: Once we are finished with Steps 1-3 we have completed one iteration step, and we compute

$$\Delta = \max \left\{ \left| (w_p^{(q)})^{\text{new}} - (w_p^{(q)})^{\text{old}} \right| \text{ with } p = 1, \dots, P_q; q = 1, \dots, Q \right\},$$

where $(w_p^{(q)})^{\text{new}}$ are the new weights just computed in step 3 and where $(w_p^{(q)})^{\text{old}}$ are the old weights used in step 1.

Interpretation: Δ measures how much the weights have changed from one iteration step to the next one.

If $\Delta > \epsilon$, where ϵ is a fixed small number e.g. $\epsilon = 0.0001$, then we return to step 1 and execute steps 1-3 again, but now we use the new weights $w_p^{(q)} = (w_p^{(q)})^{\text{new}}$ (just computed in step 3) in step 1.

If $\Delta \leq \epsilon$ then we stop and proceed as explained on the next pages.

Final Values of the Latent Variables and Path Coefficients

Once the iteration has stopped we take the weights $w_p^{(q)}$ from the last step 3, and execute step 1 once to compute the final estimates for the values of the latent variables ξ_q with these weights: in the notation from step 1

$$(\xi_{n1}, \dots, \xi_{nQ}) = \text{estimator of } \xi_1, \dots, \xi_Q \text{ for object } e_n.$$

Computation of the Path Coefficients:

The regression equations in the structural (inner) model

$$\xi_j = \sum_{\substack{q=1, \\ q: \xi_q \rightarrow \xi_j}}^Q \beta_{jq} \xi_q + \zeta_j$$

can now be solved with ordinary least squares (OLS) as we have data for ξ_j and for all ξ_q . Remember, we sum only over those indices q for which we have an arrow pointing from ξ_q to ξ_j . For notational purposes, let M_j be the set of all indices q such that there is an arrow pointing from ξ_q to ξ_j .

Computation of the Path Coefficients in the Inner Model

With $\xi_j = (\xi_{1j}, \dots, \xi_{nj})'$ and $\zeta_j = (\zeta_{1j}, \dots, \zeta_{nj})'$ the regression equations

$$\xi_{nj} = \sum_{q \text{ in } M_j} \beta_{jq} \xi_{nq} + \zeta_{nj}$$

can be written in matrix notation

$$\xi_j = \Xi_j \beta_j + \zeta_j, \quad (36)$$

where:

- The matrix Ξ_j contains the data of the latent variables ξ_q (with q in M_j) from whom an arrow points to ξ_j : More precisely the n th row of Ξ_j contains the data of object e_n for these variables.
- The vector β_j contains the coefficients that appear in (36), i.e. the coefficients β_{jq} with q in M_j .

The least squares solution to (36) is given by: $\beta_j = (\Xi_j' \Xi_j)^{-1} \Xi_j' \xi_j$

Ex. 5.1 Continued

Ex. 5.1 (d) One More Iterative Step

Using the results from Ex. 5.1 (a) to (c) for the structural equation model given in Ex. 5.1 (a), **execute a second iterative step of the iterative algorithm.**

Ex. 5.1 (e) Further Iterative Steps

Inspect the results and computations from Ex. 5.1 (a) to (d) for the structural equation model given in Ex. 5.1 (a), and in particular compare the weights $w_p^{(q)}$ from the two iterative steps and observe their effect. **Use your observations to predict the results of subsequent iterative steps. What happens after the third iterative step?**

Ex. 5.1 (f) Values of the Latent Variables and Path Coefficients

Using the results from Ex. 5.1 (a) to (d) for the structural equation model given in Ex. 5.1 (a), **stop the iterative algorithm after the third step and compute the estimates of the latent variables and the path coefficients.** Inspect your results.

Ex. 5.2: Comparing the PLS Model and the LISREL Model

Compare the coefficients of the PLS model from Ex. 5.1 with the LISREL model from Ex. 4.3. To do this, you need to consider the standardized coefficients, because the variables in the two models are scaled differently.

For a regression equation

$$X - \mu_X = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \dots + \gamma_m \xi_m + \delta,$$

where δ is the error term and $\gamma_1, \gamma_2, \dots, \gamma_m$ the coefficients, the standardized coefficients $\tilde{\gamma}_j$ are given by

$$\tilde{\gamma}_j = \gamma_j \cdot \frac{\sigma_{\xi_j}}{\sigma_X} = \gamma_j \cdot \frac{\text{standard deviation of } \xi_j}{\text{standard deviation of } X}. \quad (37)$$

For computing the standardized coefficients, we have to estimate the standard deviations in (37) by the empirical standard deviations s_{ξ_j} and s_X computed from the data.