Parameter estimation

Outline:

- **1 Types of parameters**
- **2 The criterion of estimation**
- **3 Estimation methods**
- **4 Robust rescaling**
- **5 Recommendations**
- **6 Iteration**

1 Types of parameters:

- Free parameter

  are estimated with respect to the data – direct estimation

- Restricted parameter

  the value of the parameter is set equal to the value of another parameter; the common value of the parameters is estimated – indirect estimation

- Fixed parameter

  a specific constant value is assigned to the parameter – no estimation

2 The criterion

The different **estimation methods** mainly differ with respect to the **criterion** (e.g. least-squares criterion of multiple regression)

2 The criterion

The criterion defines the way of determining the difference between the empirical matrix (**S**) and the model-implied matrix (∑) that is specified according to the vector including the parameters of the model (**θ**).

It is frequently described as function F with two entries:

$$F(\mathbf{S},\boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

2 The criterion

The criterion refers to the way of determining the difference between the empirical matrix (**S**) and the model-implied matrix (Σ) that is specified according to the vector including the parameters of the model ($\boldsymbol{\theta}$).

The difference is frequently described as function F with two entries that has to be minimized:

$$F(\mathbf{S}, \Sigma(\boldsymbol{\theta})) \rightarrow \mathbf{min}$$

2 The criterion

The criterion refers to the way of determining the difference between the empirical matrix (**S**) and the model-implied matrix (∑) that is specified according to a specific vector of parameters (**θ**).

The difference is frequently described as function F with two entries that has to be minimized:

$$F(\mathbf{S}, \Sigma(\boldsymbol{\theta})) \rightarrow \mathbf{min}$$

Estimation methods mainly differ according to the way of determining the difference

2 The criterion

…

Minimization is usually achived by means of an *iterative procedure*. This means that $\boldsymbol{\theta}$ is successively modified in such a way that the value approaches the minimum:

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) \rightarrow \mathbf{min}$$

2 The criterion

…

Minimization is usually achived by means of an iterative procedure. This means that $\boldsymbol{\theta}$ is successively modified in such a way that the value approaches the minimum:

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) \rightarrow \mathbf{min}$$

A recent development is the consideration of an additional weight matrix **W** that gives specific emphasis to some individual deviations of the matrices:

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{W}) \rightarrow \mathbf{min}$$

3 Estimation method (criterion)

- Formerly the software programs offered several estimation methods

… and the user simply selected the method that provided the best results

3 Estimation method (criterion)

- Formerly the software programs offered several estimation methods

… and the user simply selected the method that provided the best results

- Recently it is required that the selection is made according to the scale of the data (and the distribution). There are ….

  - methods for continuous data

  - methods for ordered-categorical data

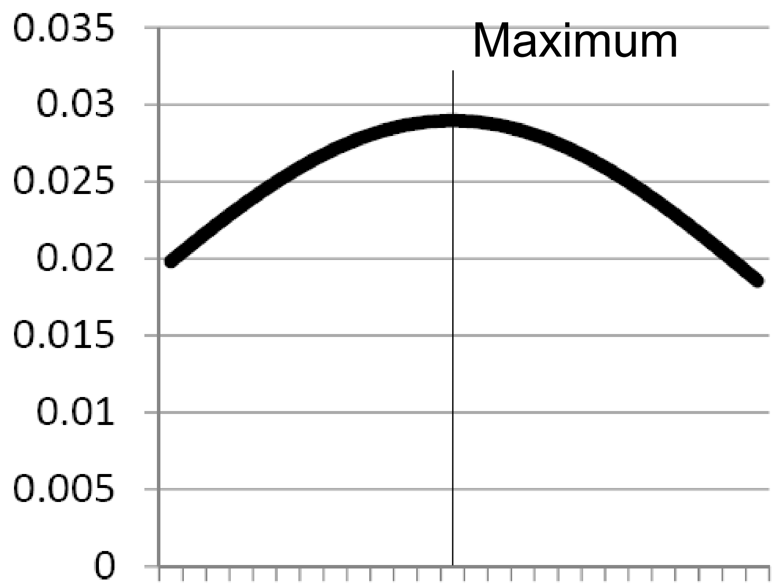
(- methods for binary data)

3 Estimation method (criterion)

Maximum likelihood method (ML):

- … returns the parameter values that are expected to show „the highest probability to underly the observed data“

Likelihood resp. F_M



… for value of θ (parameter) ($F(\mathbf{S}, \sum(\boldsymbol{\theta}))$)

3 Estimation method (criterion)

Maximum likelihood method (ML):

- … returns the parameter values that are expected to show „the highest probability to underly the observed data“
- … is the most frequently used „fitting function“

Discrepancy function (F function):

$$F_M = log|\Sigma(\theta)| - \ log|S| + tr\,[S\Sigma(\theta)^{-1}] - p$$

tr = „trace“ of a matrix

p = number of observed variables

θ : the parameters of the model

3 Estimation method (criterion)

Maximum likelihood method (ML):

Preconditions:

- … is recommended for continuous data
- … is recommended if multivariate normality is given
- … is recommended for correctly specified models
- recommended sample size: $\geq$ 200 ($\geq$ 400, $\geq$ 2000)
-

(depends on literature source and data type)

3 Estimation method (criterion)

Maximum likelihood method (ML):

Advantage:

- provides a statistic that follows the χ^2 distribution
- enables statistical testing (χ^2)
- is efficient, consistent and unbiased in combination with correct specification of the model and given preconditions

Disadvantages:

- … multivariate normality is frequently not given
- … in this case -> first-type error

3 Estimation method (criterion)

Generalized least squares (GLS)

- … is recommended for continuous data
- … requires a large sample size
- … differs from ML because of the use of the **W** matrix
- … is not used frequently

Discrepancy function:

$$F_{GLS} = \frac{1}{2} tr\left[\left(\mathbf{S} - \Sigma\left(\widehat{\boldsymbol{\theta}}\right)\right)\mathbf{W}_{GLS}{}^{-1}\right]^2$$

3 Estimation method (criterion)

Weighted Least Squares (WLS)

- … does not require multivariate normality
- … presupposes a large sample size
- … is recommended for continuous data
- … is recommended for ordered-categorical data

Discrepancy function:

$$F_{\mathrm{WLS}} = [s - \sigma(\theta)]' \mathbf{W}^{-1} [s - \sigma(\theta)]$$

with s and σ(..) as S and Σ(..) transformed into vectors

3 Estimation method (criterion)

- Advantage:
- Less sensitive to deviations from normality
- In large samples accurate χ^2 test results and standard errors

Disadvantage:

- W^{-1} matrix used in combination with WLS becomes very large for many indicators so that an investigation may no more be possible (4th derivative)
- Very large samples are necessary for achieving accurate estimates
- Problems in comparisons between models on the basis of χ^2

There are two versions

- (Simple) *Weighted Least Squares method*
- *Diagonally Weighted Least Squares method*

3 Estimation method (criterion)

Special case:

- Unweighted Least Squares (ULS)
  - Minimizes the residual matrix
  - No precondition regarding the distribution
  - No χ^2 results
  - … is seldom applied

3 Estimation method (criterion)

Application to a matrix composed of struct. data:

(20 x 20 matrix)

I1	1.016					
I2	0.107	0.349				
I3	0.102	0.081	1.002			
I4	0.136	0.072	0.180	0.616		
I5	0.172	0.100	0.174	0.149	1.026	
I6	0.085	0.062	0.107	0.110	0.069	0.460

....

....

....

3 Estimation method (criterion)

Application to a matrix composed of struct. data:

- randomly selected from a simulation study

Maximum Likelihood results

- Degrees of Freedom = 189
- Minimum Fit Function Chi-Square = 270.543 (P = 0.000)
- Normal Theory Weighted Least Squares Chi-Square = 274.332 (P = 0.000)
- Satorra-Bentler Scaled Chi-Square = 262.264 (P = 0.000335)
- Root Mean Square Error of Approximation (RMSEA) = 0.0279
- Standardized RMR = 0.0642
- Non-Normed Fit Index (NNFI) = 0.962
- Comparative Fit Index (CFI) = 0.962

3 Estimation method (criterion)

Application to a matrix composed of struct. data:

- randomly selected from a simulation study

Diagonally weighted least-square results

- Degrees of Freedom = 189
- Normal Theory Weighted Least Squares Chi-Square = 272.610 (P = 0.000)
- Satorra-Bentler Scaled Chi-Square = 257.726 (P = 0.000658)
- Chi-Square Corrected for Non-Normality = 365.061 (P = 0.00)
- Root Mean Square Error of Approximation (RMSEA) = 0.027
- Standardized RMR = 0.0617
- Non-Normed Fit Index (NNFI) = 0.964
- Comparative Fit Index (CFI) = 0.964

3 Estimation method (criterion)

Application to a matrix composed of struct. data:

- randomly selected from a simulation study

Weighted least-square results

- Degrees of Freedom = 189
- Minimum Fit Function Chi-Square = 365.061 (P = 0.00)
- ……….
- Root Mean Square Error of Approximation (RMSEA) = 0.0432
- Standardized RMR = 0.120
- Non-Normed Fit Index (NNFI) = 0.493
- Comparative Fit Index (CFI) = 0.496

Comment: robust version was not available!

4 Robust rescaling

- There are rescaling methods for the chi-square value that take deviations from normality into consideration.
- Best known and available is the method proposed by Satorra and Bentler. It starts with normal maximum likelihood estimation and subsequently modifies the chi-square value.
- The SB-scaled chi-square value is robust to non-normality.

3 Estimation method (criterion)

Application to a matrix composed of struct. data:

- randomly selected from a simulation study

Maximum Likelihood results

- Degrees of Freedom = 189
- Minimum Fit Function Chi-Square = 270.543 (P = 0.000)
- Normal Theory Weighted Least Squares Chi-Square = 274.332 (P = 0.000)
- Satorra-Bentler Scaled Chi-Square = 262.264 (P = 0.000335)
- Root Mean Square Error of Approximation (RMSEA) = 0.0279
- Standardized RMR = 0.0642
- Non-Normed Fit Index (NNFI) = 0.962
- Comparative Fit Index (CFI) = 0.962

3 Estimation method (criterion)

Application to a matrix composed of struct. data:

- randomly selected from a simulation study

Diagonally weighted least-square results

- Degrees of Freedom = 189
- Normal Theory Weighted Least Squares Chi-Square = 272.610 (P = 0.000)
- Satorra-Bentler Scaled Chi-Square = 257.726 (P = 0.000658)
- Chi-Square Corrected for Non-Normality = 365.061 (P = 0.00)
- Root Mean Square Error of Approximation (RMSEA) = 0.027
- Standardized RMR = 0.0617
- Non-Normed Fit Index (NNFI) = 0.964
- Comparative Fit Index (CFI) = 0.964

5 Recommendations

- If possible, select ML!
- In other cases, follow the recommendations by Finney, DiStefano and Kopp (2016) (see next Table)

TYPE OF DATA	Suggestions	Caveats/Notes
Continuous Data		
Approximately Normally Distributed	•Use ML estimation	•The assumptions of ML are met and estimates should be unbiased, efficient, and consistent
Moderately Nonnormal	•Use S-B/robust method to correct χ^2 and standard errors for slight nonnormality	•Given the availability of S-B/robust methods in the software packages, one could employ and report findings from both ML estimation and S-B/robust method
(skew < 2, kurtosis < 7)	•Use ML estimation, as it is fairly robust to these conditions	
Severely Nonnormal (skew > 2, kurtosis > 7)	•Use S-B/robust method	•S-B performs best, although may suffer from loss of χ^2 power

TYPE OF DATA	Suggestions	Caveats/Notes
Ordered Categorical Data		
Number of Ordered Categories is 6 or more	•Treat data as continuous and use S-B scaling methods with ML estimation	•Parameter estimates from S-B/robust approach equal ML-based estimates implying that they will be attenuated to some extent
	•Treat data as categorical and use robust DWLS estimator	•Robust DWLS will adjust the parameter estimates, standard errors, and fit indices for the categorical nature of the data
		•One could employ and report findings from both estimation methods
Number of Ordered Categories is 5 or less	•Treat data as categorical and use robust DWLS estimator	•Robust DWLS will adjust the parameter estimates, standard errors, and fit indices for the categorical nature of the data
	•	•

6 Iteration

The parameter estimation occurs using the …

… Expectation Maximization Algorithmus (EM) - or – special case:

… Newton Raphson Algorithmus

that are iterativ

- The iteration starts with **starting values** as parameter values giving rise to the model-implied matrix
- The fitting function yields a fit result for the starting values
- The *expectation* component of the algorithm prospects possible improvement

6 Iteration

The parameter estimation occurs using the …

….

- The *maximization* component of the algorithm revises the parameter values
- The fitting function yields a new fit result
- The new and old fit results are compared

… in the case of an improvement in fit ….

- The *expectation* component of the algorithm again prospects possible improvment ……………………..

…………………< the iteration goes on >……………

6 Iteration

The parameter estimation occurs using the …

….

- The *maximization* component of the algorithm revises the parameter values
- The fitting function yields a new fit result
- The new and old fit results are compared

… in the case of an improvement in fit ….

… in the case of no improvement or reaching the maximum number of iterations …

End of iteration!

In sum

- Chose the estimation method in considering the data!
- … what is the scale
- … what is the distribution

QUESTIONS REGARDING COURSE UNIT 10

- What are the preconditions of the maximum likelihood method?
- Which statistic is provided that enables the estimation of the error probability?
- What is the reason for selecting a robust estimation method?

Literature:

Finney, S. J., DiStefano, C., & Kopp, J. P. (2016). Overview of estimation methods and preconditions for their application with structural equation modeling. In K. Schweizer, & C. DiStefano (Eds.), *Principles and Methods of Test Construction* (pp. 136-165). Göttingen: Hogrefe Publishing.

Kline, R. B. (2011). *Principles and practices of structural equation modeling* (3rd edition) (Chapter 1: Introduction). New York, NY: The Guilford Press.