EVALUATION OF STRUCTURAL EQUATION MODELS

I. Issues related to the initial specification of theoretical models of interest

1. Model specification:

    a. Measurement model:

        (i) EFA vs. CFA
        (ii) reflective vs. formative indicators [see Appendix A]
        (iii) number of indicators per construct [see Appendix B]

        - total aggregation model
        - partial aggregation model
        - total disaggregation model

    b. Latent variable model:

        (i) recursive vs. nonrecursive models
        (ii) alternatives to the target model [see Appendix C for an example]

2. Model misspecification

    a. omission/inclusion of (ir)relevant variables
    b. omission/inclusion of (ir)relevant relationships
    c. misspecification of the functional form of relationships

3. Model identification

4. Sample size

5. Statistical assumptions

II. Data screening

1. Inspection of the raw data

    a. detection of coding errors
    b. recoding of variables
    c. treatment of missing values

2. Outlier detection

3. Assessment of normality

4. Measures of association

    a. regular vs. specialized measures
    b. covariances vs. correlations
    c. non-positive definite input matrices

III. Model estimation and testing

1. Model estimation

2. Estimation problems

    a. nonconvergence or convergence to a local optimum
    b. improper solutions
    c. problems with standard errors
    d. empirical underidentification

3. Overall fit assessment [see Appendix D]

4. Local fit measures
   [see Appendix E on how to obtain robust standard errors]

   a. Measurement model

      (i) factor loadings, factor (co)variances, and error variances
      (ii) reliabilities and discriminant validity

   b. Latent variable model

      (i) structural coefficients and equation disturbances
      (ii) direct, indirect, and total effects [see Appendix F]
      (iii) explained variation in endogenous constructs

5. Power [see Appendix G]

6. Model modification and model comparison [see Appendix H]

   a. Measurement model
   b. Latent variable model

7. Model-based residual analysis

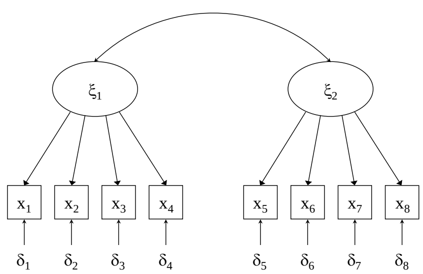
8. Cross-validation

9. Model equivalence and near equivalence [see Appendix I]

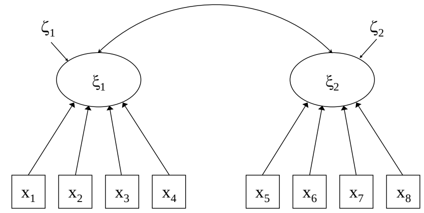
10. Latent variable scores [see Appendix J]

Appendix A: Reflective vs. formative indicators

Reflective indicators



Formative indicators



Criteria for distinguishing between reflective and formative indicator models

- Are the indicators manifestations of the underlying construct or defining characteristics of it?
- Are the indicators conceptually interchangeable?
- Are the indicators expected to covary?
- Are all of the indicators expected to have the same antecedents and/or consequences?

Based on MacKenzie, Podsakoff and Jarvis, JAP 2005, pp. 710-730.

Appendix B: Number of indicators per factor

Total aggregation model:

(a) If measurement error in indicator i is ignored, then

$$\lambda_{ij} = 1 \ \ and \ \ \theta_{ii} = 0$$

(b) If indicator i is treated as fallible but reliability is unknown, a small amount of measurement error might be assumed:

$$\lambda_{ij} = 1 \ \ and \ \ \theta_{ii} = .1 s_{x_i}^2$$

(c) If indicator i is treated as fallible and an estimate of reliability (ρ_{ii}) is available:

$$\lambda_{ij} = 1 \ \ and \ \ \theta_{ii} = (1 - \rho_{ii}) s_{x_i}^2$$

Partial aggregation model:

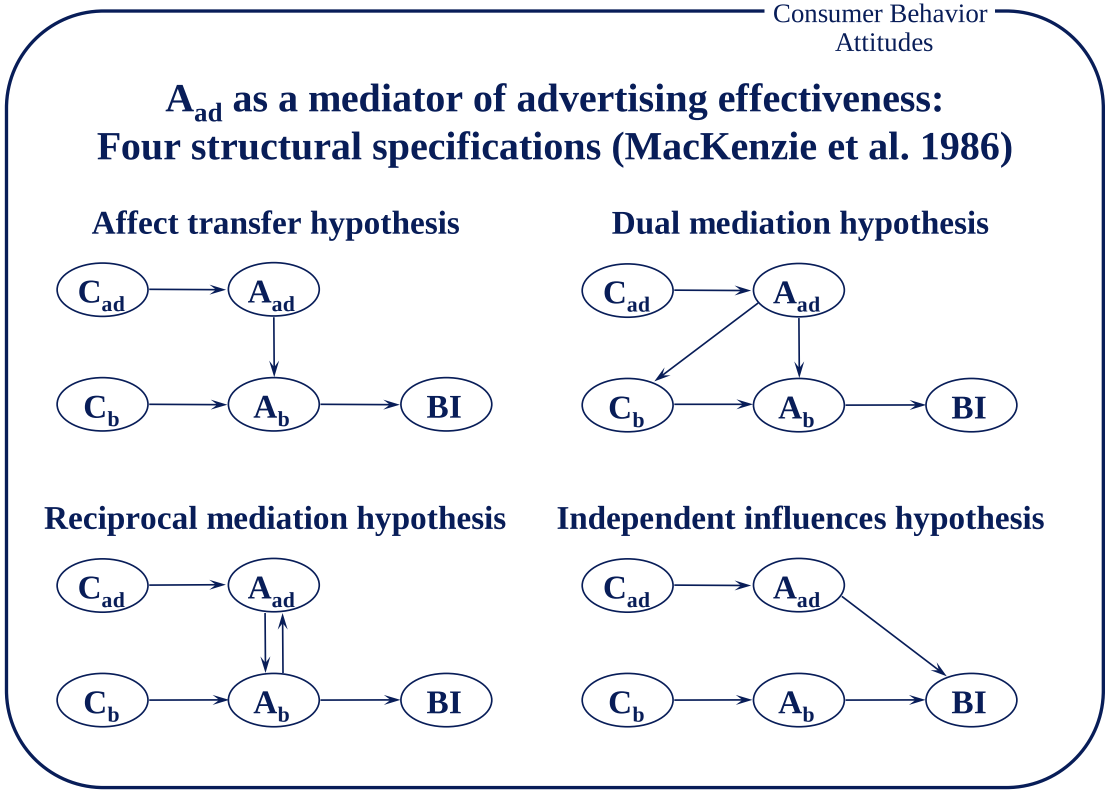
Recommendations for item parceling (based on Holt 2004):

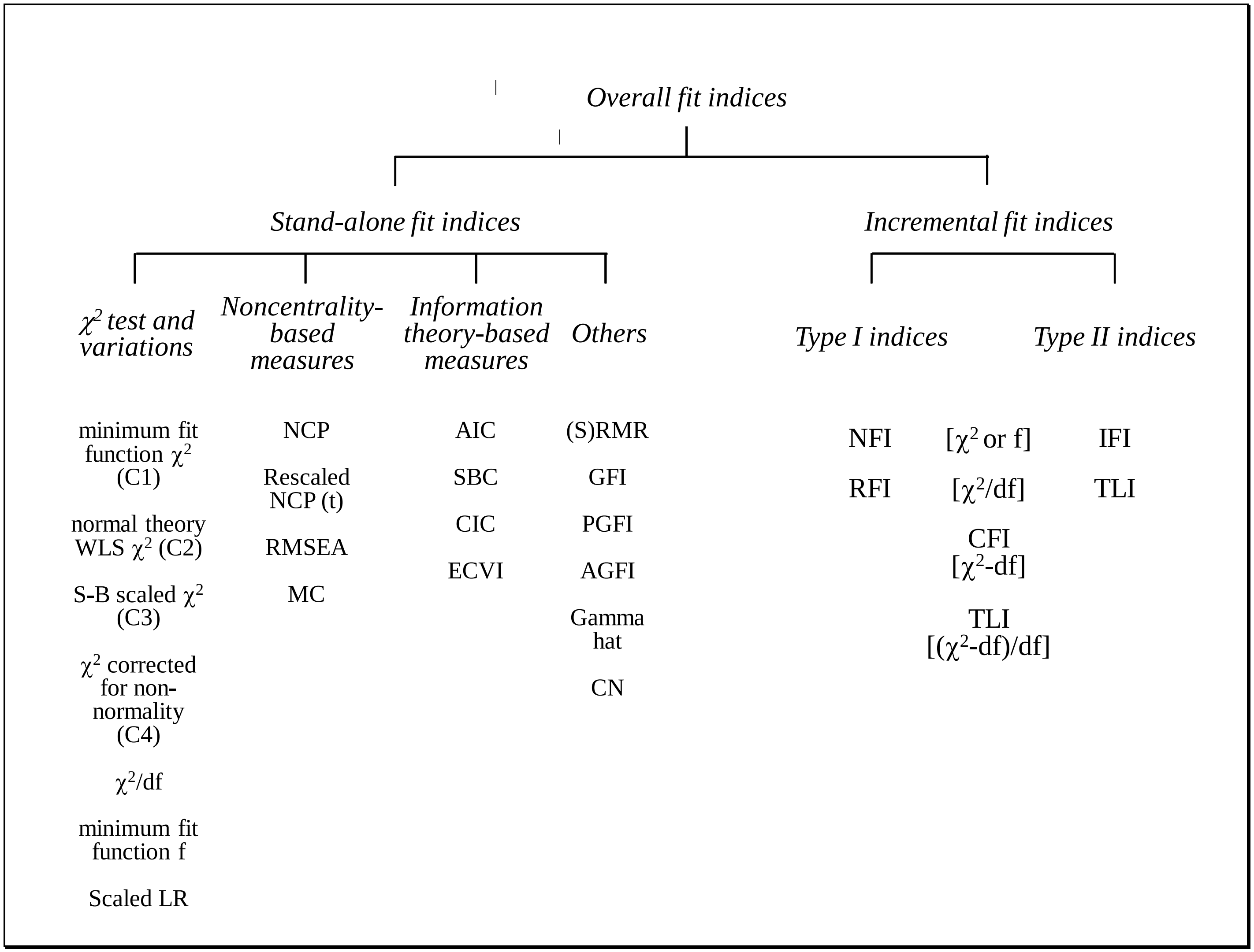
- if a factor is unidimensional, the items may be parceled randomly;
- if a factor is multidimensional, item parcels should be based on similar facets of the construct;
- item parceling may be used strategically to correct for non-normality problems;
- if the factor structure is not well-understood, do not parcel;

Total disaggregation model:

in principle, more indicators are better than fewer, but if the number of indicators is too large, it will be difficult to achieve a good model fit;

Appendix C: Specification of a priori models to be compared using SEM




Overall fit indices
Stand-alone fit indices
Incremental fit indices
χ2 test and variations
Noncentrality-based measures
Information theory-based measures
Others
Type I indices
Type II indices
minimum fit function χ2 (C1)
normal theory WLS χ2 (C2)
S-B scaled χ2 (C3)
χ2 corrected for non-normality (C4)
χ2/df
minimum fit function f
Scaled LR
NCP
Rescaled NCP (t)
RMSEA
MC
AIC
SBC
CIC
ECVI
(S)RMR
GFI
PGFI
AGFI
Gamma hat
CN
NFI
RFI
[χ2 or f]
[χ2/df]
CFI
[χ2-df]
TLI
[(χ2-df)/df]
IFI
TLI

Summary of overall fit (or lack-of-fit) indices

(a) stand-alone fit indices:

Index	*Definition of the index*[1]	*Characteristics*[2]	*Interpretation and use of the index*
Minimum fit function chi-square (C1)	(N-1)f	BF, SA, NN, NP	tests the hypothesis that the specified model provides a perfect fit (within the limits of sampling error); the obtained χ^2 value should be smaller than χ^2_{crit}; note that different discrepancy functions will yield different χ^2 values;
Normal theory weighted least squares (WLS) chi-square (C2)	χ^2	BF, SA, NN, NP	same interpretation as C1 and asymptotically equivalent to it;
Satorra-Bentler scaled (robust) chi-square (C3)	χ^2_{robust}	BF, SA, NN, NP	scaled χ^2 statistic corrected for violations of multivariate normality; same interpretation as C1;
Chi-square corrected for non-normality (Browne 1984; C4)	χ^2_{nn}	BF, SA, NN, NP	χ^2 statistic corrected for violations of multivariate normality; same interpretation as C1;
Elliptical corrected χ^2 (Browne 1982; χ^2_{ELL})	$\chi^2/(1+\kappa)$	BF, SA, NN, NP	χ^2 when the variables have a multivariate elliptical distribution; the multivariate normal distribution is a special case with $\kappa=0$; same interpretation as C1;
Chi-square over degrees of freedom ratio (χ^2/df)	χ^2/df	BF, SA, NN, P	estimate of how many times larger the obtained χ^2 value is than the expected value (under H_0) of a χ^2 variate with a given number of df; ratios smaller than 2 or 3 are sometimes considered desirable;
Minimum fit function value or fit criterion (f)	f	BF, SA, NN, NP	minimum of the discrepancy function comparing S with $\hat{\Sigma}$; if the fit of the model is perfect (i.e., S= $\hat{\Sigma}$), f will be zero; can be used to compare competing models;
Scaled likelihood ratio (Marsh et al. 1988; LHR)	exp [-(1/2)f]	GF, SA, N, NP	f scaled to fall within the [0, 1] interval; the closer to one the better the fit;

Noncentrality parameter (Browne and Cudeck 1993; NCP)	$(\chi^2\text{-df})$	BF, SA, NN, NP	estimate of how well the fitted model approximates the population covariance matrix; can be interpreted as a weighted sum of squares of discrepancies between the parameters of the approximating model and the parameters of the model yielding Σ; a confidence interval around NCP is available;
Rescaled NCP (McDonald and Marsh 1990) or population discrepancy function value (t, F_0)	$(\chi^2\text{-df})/(N\text{-}1)$	BF, SA, NN, NP	estimate of error due to approximation; a confidence interval around t is available;
Root mean squared error of approximation (Browne and Cudeck 1993; Steiger 1990; RMSEA)	$(t/df)^{½}$	BF, SA, NN, P	estimate of how well the fitted model approximates the population covariance matrix per df; Browne and Cudeck (1993) suggest that a value of .05 indicates a close fit and that values up to .08 are reasonable; Hu and Bentler (1999) recommend a cutoff value of .06; a p-value for testing the hypothesis that the discrepancy is smaller than .05 may be calculated (test of close fit);
McDonald’s (1989) measure of centrality (MC)	exp [-(1/2)t]	GF, SA, AN, NP	t scaled to fall within the [0, 1] interval; the closer to one the better the fit; Hu and Bentler (1999) recommend a cutoff value of close to .9;
Akaike’s (1987) information criterion (AIC)	$[\chi^2+2r]$ or $[\chi^2–2df]$	BF, SA, NN, P	based on statistical information theory and used for testing competing models; the model with the smallest AIC is selected;
Schwarz’s (1978) Bayesian criterion (SBC)	$[\chi^2+r \ln N]$ or $[\chi^2–df \ln N]$	BF, SA, NN, P	modification of AIC; same interpretation and use, although it may lead to a different rank order of models;
Bozdogan’s (1987) consistent information criterion (CIC)	$[\chi^2+r(\ln N+1)]$ or $[\chi^2–df(\ln N+1)]$	BF, SA, NN, P	modification of AIC; same interpretation and use, although it may lead to a different rank order of models;
Expected cross-validation index (ECVI; Browne and Cudeck 1989)	[f+(2r)/(N–1)] or [f+(2r)/(N–(p+q) –2)]	BF, SA, NN, P	single-sample approximation to the cross-validation coefficient obtained from a validation sample; similar to AIC and same rank order; a confidence interval can be calculated to take into account the precision of the estimate; not recommended by Hu and Bentler (1999);
Root mean residual (Jöreskog and Sörbom 1993; (S)RMR)	$\{2\sum\sum(s_{ij} - \hat{\sigma}_{ij})^2/ [(p+q)(p+q+1)]\}^{1/2}$	BF, SA, N or NN, NP	measure of the average size of residuals between the fitted and sample covariance matrices; if a correlation matrix is analyzed, RMR is “standardized” to fall within the [0, 1] interval (SRMR), otherwise it is only bounded from below; a cutoff of .05 is often used for SRMR; Hu and Bentler (1999) recommend a cutoff value close to .08;

Goodness-of-fit index (Jöreskog and Sörbom 1993; GFI)	$1-\{f[S, \hat{\Sigma}]/f[S, \hat{\Sigma}(0)]\}$	GF, SA, AN, NP	proportion of the variances and covariances in S accounted for by the fitted model; should fall between 0 and 1, although it can be smaller than 0; values greater than .9 are sometimes deemed desirable; not recommended by Hu and Bentler (1999);
Parsimonious goodness-of-fit index (PGFI; Mulaik et al. 1989)	$2df/[(p+q)(p+q+1)]\cdot GFI$ or $[df_t/df_n]\cdot GFI$	GF, SA, AN, P	GFI adjusted for (loss of) df (by multiplying by the parsimony index); can be used to compare competing models;
Adjusted goodness-of-fit index (Jöreskog and Sörbom 1993; AGFI)	$1-[(p+q)(p+q+1)/2df]\cdot(1-GFI)$	GF, SA, AN, P	GFI adjusted for (loss of) df (adjustment based on the same logic as in the case of an adjusted R^2); values greater than .9 are sometimes deemed desirable; not recommended by Hu and Bentler (1999);
Gamma hat (Steiger 1989)	$(p+q)/(p+q+2t)$	GF, SA, AN, NP	population equivalent of GFI expressed in terms of the non-centrality parameter; Hu and Bentler (1999) recommend a cutoff value of .95;
Hoelter's (1983) critical N (CN)	$[\chi^2_{crit}/f+1]$ or $\{[Z_{crit}+(2df-1)^{1/2}]^2/[2\chi^2/(N-1)]\}+1$	GF, SA, NN, NP	estimate of the sample size at which the null hypothesis that the model fits is rejected; Hoelter suggests that CN should be larger than 200; not recommended by Hu and Bentler (1999);

(a) incremental fit indices:

Bentler and Bonett's (1980) normed fit index (NFI) or Δ_1	$1-[\chi^2_t/\chi^2_n]$	GF, IM, N, NP	measure of the proportionate improvement in fit (defined in terms of f or χ^2) as one moves from the baseline to the target model; values greater than .9 are usually deemed desirable; problem that it is biased downward for small N; not recommended by Hu and Bentler (1999);
Parsimonious normed fit index (James et al. 1982; PNFI)	$df_t/df_n \cdot [1-\chi^2_t/\chi^2_n]$	GF, IM, N, P	NFI adjusted for (loss of) df (by multiplying by the parsimony index); can be used to compare competing models;
Bollen's (1989) nonnormed fit index Δ_2 or incremental fit index (IFI)	$[\chi^2_n-\chi^2_t]/[\chi^2_n-df_t]$	GF, IM, AN, P	modified version of NFI designed to lessen its dependence on sample size; however, it may be biased upward for small N when the model is misspecified, and the parsimony correction may be inappropriate; Hu and Bentler (1999) recommend a cutoff value of .95;
Bollen's (1986) "normed" index ρ_1 or relative fit index (RFI)	$1-[(\chi^2_t/df_t)/(\chi^2_n/df_n)]$	GF, IM, AN, P	similar to NFI, except that the fit of the target model is compared to the fit of the baseline model *per df*; problem that it is biased downward for small N; not recommended by Hu and Bentler (1999);
Bentler's (1990) normed Comparative Fit Index (CFI)	$1-[\max(\chi^2_t-df_t,0)/\max(\chi^2_n-df_n,\chi^2_t-df_t,0)]$	GF, IM, N, NP	measure of the proportionate improvement in fit (defined in terms of noncentrality) as one moves from the baseline to the target model; Hu and Bentler (1999) recommend a cutoff value of .95;
McDonald and Marsh's (1990) nonnormed relative noncentrality index (RNI)	$1-[(\chi^2_t-df_t)/(\chi^2_n-df_n)]$	GF, IM, AN, NP	similar to CFI in most cases, but not restricted to fall strictly within the [0, 1] interval; Hu and Bentler (1999) recommend a cutoff value of .95;
Tucker and Lewis (1973) nonnormed fit index (TLI, NNFI) or ρ_2	$[\chi^2_n/df_n-\chi^2_t/df_t]/[\chi^2_n/df_n-1] = 1-[(t_t/df_t)/(t_n/df_n)]$	GF, IM, AN, P	measure of the proportionate improvement in fit (defined in terms of noncentrality) as one moves from the baseline to the target model, per df; Hu and Bentler (1999) recommend a cutoff value of .95;

(footnotes on next page)

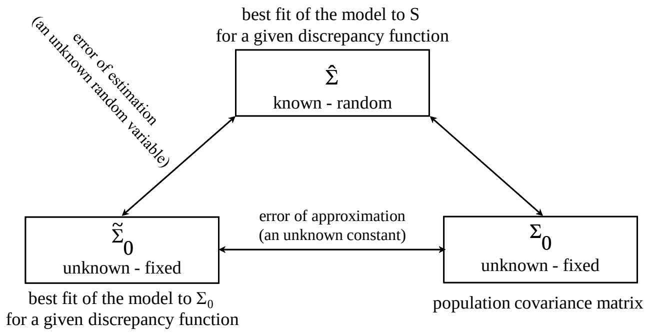
[1] N = sample size;
df = degrees of freedom;
r = number of parameters estimated;
(p+q) = number of observed variables;
κ = Mardia-based kappa, an estimate of multivariate kurtosis (Browne 1982);
χ^2_{crit} = critical value of the χ^2 distribution with the appropriate number of degrees of freedom and for a given significance level;
Z_{crit} = critical value of the normal distribution for a given level of statistical significance.

The subscripts n and t refer to the null (or baseline) and target models, respectively.

In the definition of GFI, $f[S, \hat{\Sigma}]$ refers to the minimum of the fitting function after the target model has been fitted and $f[S, \hat{\Sigma}(0)]$ refers to the minimum of the fitting function before any model has been fitted.

[2] GF = goodness-of-fit index;
BF = badness-of-fit index;
SA = stand-alone fit index;
IM = incremental fit index;
N = normed (in the sample) fit index;
AN = normed (in the population) fit index, but only approximately normed in the sample (i.e., can fall outside the [0, 1] interval);
NN = nonnormed fit index;
NP = no correction for parsimony;
P = correction for parsimony.

Types of error in covariance structure modeling



Incremental fit indices

- type I indices: $\frac{GF_t - GF_n}{GF_t}$ *or* $\frac{BF_n - BF_t}{BF_n}$

- type II indices: $\frac{GF_t - GF_n}{E(GF_t) - GF_n}$ *or* $\frac{BF_n - BF_t}{BF_n - E(BF_t)}$

where: GF_t, BF_t = value of some stand-alone index for the target model;

GF_n, BF_n = value of the stand-alone index for the null model;

$E(GF_t)$, $E(BF_t)$ = expected value of GF_t or BF_t assuming that the target model is true;

APPENDIX E: HOW TO OBTAIN ROBUST MODEL STATISTICS

PRELIS COMMAND FILE (CFA-ROBUST-PR2):

```
SCREENING OF COUPON OBSERVATIONS
DA NI=9
LA
ID AA1T1 AA2T1 AA3T1 AA4T1 AA1T2 AA2T2 AA3T2 AA4T2
RA=D:\M554\EDEN3\CFA.DAT
CO ALL
OU MA=CM AC=CFA.ACC
```

SIMPLIS COMMAND FILE (CFA3.SPL):

```
TITLE
CONFIRMATORY FACTOR MODEL (ATTITUDE TOWARD USING COUPONS MEASURED AT TWO POINTS IN TIME)
OBSERVED VARIABLES
ID AA1T1 AA2T1 AA3T1 AA4T1 AA1T2 AA2T2 AA3T2 AA4T2
RAW DATA FROM FILE=D:\M554\EDEN3\CFA.DAT
ASYMPTOTIC COVARIANCE MATRIX FROM FILE=D:\M554\EDEN3\CFA.ACC
LATENT VARIABLES
AAT1 AAT2
SAMPLE SIZE 250
RELATIONSHIPS
AA1T1 AA2T1 AA3T1 AA4T1 = AAT1
AA1T2 AA2T2 AA3T2 AA4T2 = AAT2
SET THE VARIANCE OF AAT1 AAT2 TO 1
SET THE ERROR COVARIANCE OF AA1T1 AND AA1T2 FREE
SET THE ERROR COVARIANCE OF AA2T1 AND AA2T2 FREE
SET THE ERROR COVARIANCE OF AA3T1 AND AA3T2 FREE
SET THE ERROR COVARIANCE OF AA4T1 AND AA4T2 FREE
OPTIONS SC RS MI WP
PATH DIAGRAM
END OF PROBLEM
```

LISREL COMMAND FILE (CFA.LS8):

```
Confirmatory factor model
DA NI=9 NO=0
LA
id aa1t1 aa2t1 aa3t1 aa4t1 aa1t2 aa2t2 aa3t2 aa4t2
ra fi=d:\m554\eden3\cfa.dat
ac fi=d:\m554\eden3\cfa.acc
se
2 3 4 5 6 7 8 9/
MO nx=8 nk=2 ph=st td=sy,fr
fr lx 1 1 lx 2 1 lx 3 1 lx 4 1 lx 5 2 lx 6 2 lx 7 2 lx 8 2
pa td
1
0 1
0 0 1
0 0 0 1
1 0 0 0 1
0 1 0 0 0 1
0 0 1 0 0 0 1
0 0 0 1 0 0 0 1
lk
aat1 aat2
ou ml sc
```

Overall goodness-of-fit statistics:

Degrees of Freedom = 15
Minimum Fit Function Chi-Square = 26.76 (P = 0.031)
Normal Theory Weighted Least Squares Chi-Square = 26.28 (P = 0.035)
Satorra-Bentler Scaled Chi-Square = 15.34 (P = 0.43)
Chi-Square Corrected for Non-Normality = 13.86 (P = 0.54)

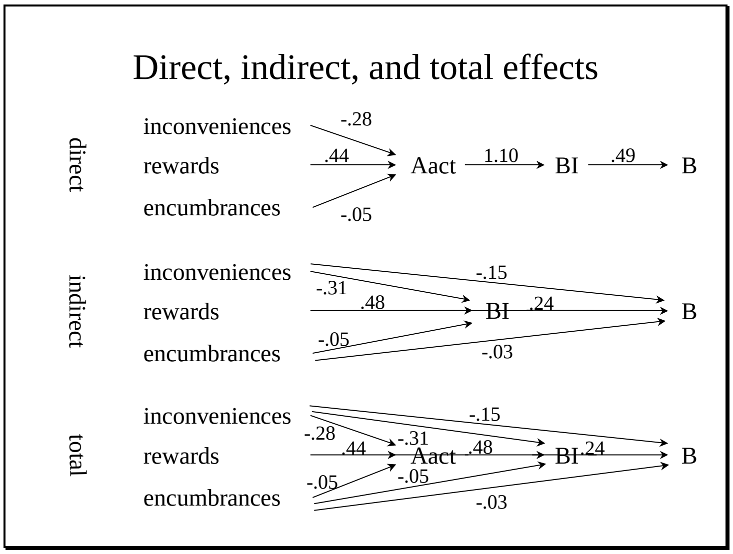
Local fit indices:

construct	*parameter*	*parameter estimate (ML)*	*estimated standard error (ML)*	*estimated robust standard error*
Aact(t1)				
	λ^x_{11}	1.084	.073	.077
	λ^x_{21}	1.108	.069	.077
	λ^x_{31}	0.923	.071	.081
	λ^x_{41}	1.221	.075	.078
	θ^δ_{11}	0.665	.073	.110
	θ^δ_{22}	0.483	.060	.116
	θ^δ_{33}	0.755	.076	.101
	θ^δ_{44}	0.546	.069	.128
Aact(t2)				
	λ^x_{52}	1.188	.073	.073
	λ^x_{62}	1.170	.068	.077
	λ^x_{72}	0.976	.066	.073
	λ^x_{82}	1.236	.073	.077
	θ^δ_{55}	0.563	.064	.093
	θ^δ_{66}	0.395	.051	.094
	θ^δ_{77}	0.555	.058	.075
	θ^δ_{88}	0.480	.060	.109
φ_{21}		.885	.020	.027

Note: Error covariances not shown.

Appendix F:

Direct, indirect and total effects



Note: To get indirect and total effects in LISREL, use the EF keyword on the OU line.

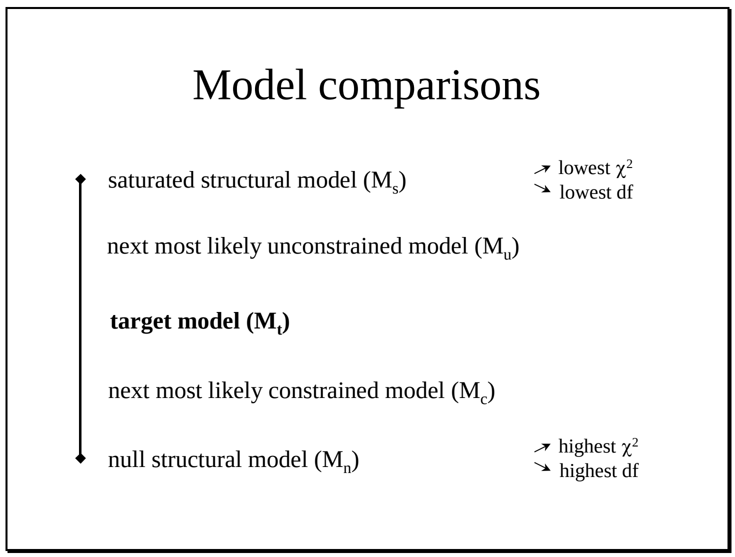
Appendix G: Power

Decision	True state of nature: H_0 true	True state of nature: H_0 false
Accept H_0	Correct decision	Type II error (β)
Reject H_0	Type I error (α)	Correct decision

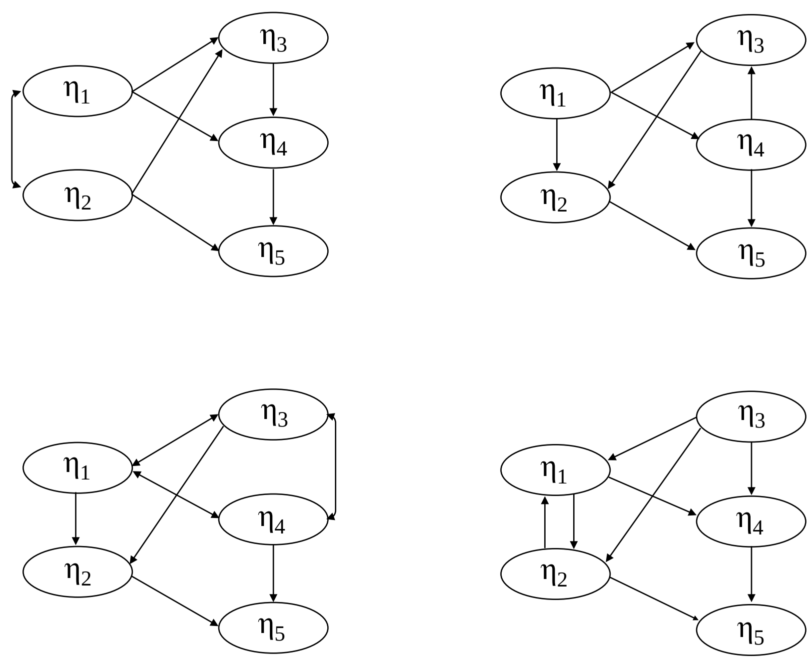
test statistic	power: low	power: high
non-significant		
significant		

Appendix H:

Model comparisons



Appendix I:
Four equivalent models (Kline 2005)



Appendix J:
Latent variable scores

Sometimes it is of interest to get estimated latent variable scores for η and ξ (e.g., if one wants to estimate nonlinear relationships among the latent variables) and estimated observational residuals. To obtain these variables, do the following:

Step 1: Use the following Prelis command file (cfa1.pr2) to create a psf file (which will be called cfa.psf)

```
Getting latent variable scores for coupon data - creating a psf file
DA NI=8
LA
aa1t1 aa2t1 aa3t1 aa4t1 aa1t2 aa2t2 aa3t2 aa4t2
RA=d:\m554\cfa\LatVarScores\cfa.dat
CO ALL
OU ra=cfa.psf
```

Step 2: Use the following Simplis file (cfa2.spl) to create a new psf file (which will automatically be called cfanew.psf)

```
Raw data from file cfa.psf
Latent Variables: AAT1 AAT2
Relationships
aa1t1 aa2t1 aa3t1 aa4t1 = AAT1
aa1t2 aa2t2 aa3t2 aa4t2 = AAT2
Set the Variance of AAT1 AAT2 to 1
Path Diagram
PSFfile cfa.psf
Estimated Residuals
End of Problem
```

The file cfanew.psf will contain both the latent variable scores for AAT1 and AAT2, as well as observational residuals for all observed indicators. You can read this file into Prelis/Lisrel and output the data to another file for further analysis in SAS, etc.

Alternatively, you can use the following Prelis file (called cfa3.pr2), which will output the latent variable scores to a text file called cfa3.fsc (the file cfa2.msf is a system file that was created in Step 2):

```
Getting latent variable scores for coupon data
DA NI=8
LA
aa1t1 aa2t1 aa3t1 aa4t1 aa1t2 aa2t2 aa3t2 aa4t2
RA=d:\m554\cfa\LatVarScores\cfa.dat
CO ALL
FS=d:\m554\cfa\LatVarScores\cfa2.msf
OU ma=cm
```