General structural model – Part 1: Power of testing, mean-structure, etc.

Psychology 588: Covariance structure and factor models

Estimation

- Fitting functions (F_{ML}, F_{GLS}, F_{ULS}) of a general SE model have the same forms as those for path modeling only of observed variables and CFA, but the implied covariance matrix is differently defined, e.g.,

$$F_{\mathrm{ML}} = \log\left|\hat{\mathbf{\Sigma}}\right| + \mathrm{tr}\left(\mathbf{S}\hat{\mathbf{\Sigma}}^{-1}\right) - \log\left|\mathbf{S}\right| - (p+q)$$

- Properties of the ML, GLS and ULS estimators hold essentially the same

- Given a converged solution, all estimates must be <u>substantively</u> sensible --- Exercise: fit the model explained in p. 334 to the political democracy data (with and without the equal-loading constraints in nested modeling approach); which are given in the data directory (poldemcov.xls)

Power of chi-square tests

- Given a pair of nesting-nested models, we can set up H_0 and H_{a} as follows:

  H_0 --- the constraints that make the only difference between the two models are correct, such that $\boldsymbol{\theta}' = [\boldsymbol{\theta}'_{\mathrm{a}}, \boldsymbol{\theta}'_{\mathrm{b}}]$, $\boldsymbol{\theta}_{\mathrm{a}} = \boldsymbol{\theta}_0$, and $\boldsymbol{\theta}_{\mathrm{b}}$ contains free parameters for both H_0 and H_{a}

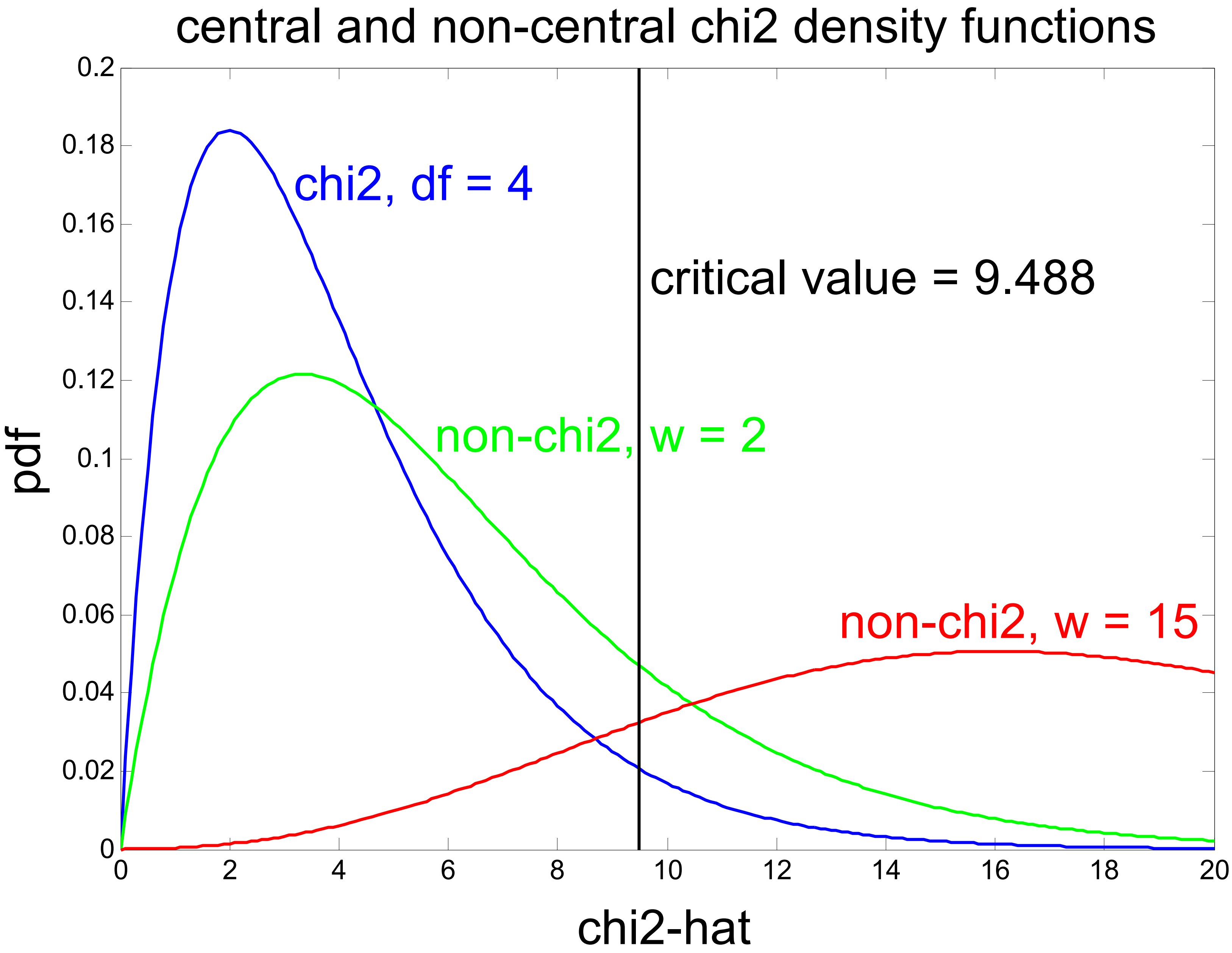
  H_{a} --- $\boldsymbol{\theta}_{\mathrm{a}} \neq \boldsymbol{\theta}_0$

- The constraints are often $\boldsymbol{\theta}_{\mathrm{a}} = \mathbf{0}$, though not necessary; it equally holds for constraints at nonzero constants

- If the nesting model (i.e., H_{a} is true) has 0 model df, the test is about goodness of fit of a hypothesized model

- Type-I error occurs only when H_0 is true while Type-II error (and hence power) is relevant only when H_a is true
  - Nominal vs. true Type-I error rate --- do we know true Type-I error rate in practice? And true power?
- All chi-square tests so far assumed true H_0
- When H_a is true, the chi-square values computed under H_0 do not follow the χ^2 distribution we use for null-hypothesis testing (which is called "central" χ^2 distribution); instead, they follow noncentral χ^2 distribution, which has one more parameter, <u>noncentrality</u> that depends on true values of $\boldsymbol{\theta}_\text{a}$
- Thus, calculating power of a chi-square test boils down to estimation of the noncentrality parameter

- Noncentrality, ω essentially defines how much chi-square values deviate from their incorrect expectation ($= df$) due to the wrong assumption of true H_0 --- i.e., $E\left(\chi^2_{\text{non}}\right) = df + \omega$
  - As in usual null hypothesis testing, smaller α leads to smaller Type I error and power; and "small effects" are hard to detect and so resulting in weak tests
  - What's wrong with Fig. 8.6 (p. 339)?



central and non-central chi2 density functions
pdf
chi2-hat
chi2, df = 4
critical value = 9.488
non-chi2, w = 2
non-chi2, w = 15
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
2
4
6
8
10
12
14
16
18
20

- The Wald statistic is defined under true H_0, i.e., it's chi-square distributed when $\boldsymbol{\theta}_\mathrm{a} = \boldsymbol{\theta}_0$ with $df = \#(\boldsymbol{\theta}_\mathrm{a})$

$$W = \left(\hat{\boldsymbol{\theta}}_a - \boldsymbol{\theta}_0\right)' \operatorname{acov}\left(\hat{\boldsymbol{\theta}}_a\right)^{-1} \left(\hat{\boldsymbol{\theta}}_a - \boldsymbol{\theta}_0\right)$$

where $\operatorname{acov}\left(\hat{\boldsymbol{\theta}}_a\right)$ is an estimate of asymptotic covariance matrix of the parameters $\boldsymbol{\theta}_\mathrm{a}$

- Under true H_a, the noncentrality parameter ω is defined as:

$$\omega = \left(\boldsymbol{\theta}_a - \boldsymbol{\theta}_0\right)' \operatorname{acov}\left(\hat{\boldsymbol{\theta}}_a\right)^{-1} \left(\boldsymbol{\theta}_a - \boldsymbol{\theta}_0\right)$$

where $\boldsymbol{\theta}_\mathrm{a}$ is the true parameter values (either empirically obtained or rationally specified) and $\boldsymbol{\theta}_0$ is constrained values

How to compute noncentrality?

Based on the *W* statistic,

1. Determine $\boldsymbol{\theta}' = [\boldsymbol{\theta}'_{\mathrm{a}}, \boldsymbol{\theta}'_{\mathrm{b}}]$ (e.g., user-provided plausible values for $\boldsymbol{\theta}_{\mathrm{a}}$) --- fixing $\boldsymbol{\theta}_{\mathrm{a}}$ at particular values amounts to setting a particular "effect size", though the size itself is not apparent

2. Generate model-implied matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ according these constants (or effect)

3. Fit the model under H_{a} (with $\boldsymbol{\theta}_{\mathrm{a}}$ as free parameters) to $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ so as to obtain $\mathrm{acov}\left(\hat{\boldsymbol{\theta}}_{\mathrm{a}}\right)$ the sub-matrix of $\mathrm{acov}\left(\hat{\boldsymbol{\theta}}\right)$,

4. Plug in $\mathrm{acov}\left(\hat{\boldsymbol{\theta}}_{\mathrm{a}}\right)$ and $(\boldsymbol{\theta}_{\mathrm{a}} - \boldsymbol{\theta}_0)$ into the formula

- Given an estimate of ω (and df, cut-off χ^2 value at α), we can calculate power by mapping on the corresponding noncentral χ^2 distribution (available in computer programs, e.g., MATLAB)

- If only one parameter is considered and the common parameters $\boldsymbol{\theta}_{\mathrm{b}}$ and θ_{a} are to be set at their estimates under H_{a}, then the asymptotic variance of θ_{a} is simply square of the standard error of $\hat{\theta}_{\mathrm{a}}$ available in usual SEM output, and so ω is readily available as: $\omega = \left(\hat{\theta}_{\mathrm{a}} - \theta_0\right)^2 \Big/ \sigma^2_{\theta_a}$

- There are cases where the fitting of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ under H_{a} is not feasible (e.g., the alternative model is not identifiable, or one of several equivalent models so that it's not unique) --- in such cases, the LR approach may be used

  Note -- if the alternative model is not unique, $\mathrm{acov}\left(\hat{\boldsymbol{\theta}}_{\mathrm{a}}\right)$ is not meaningful

Based on the LR statistic,

1. & 2. Do the same as before

3. Fit the model under H_0, i.e., with the constraints of $\boldsymbol{\theta}_{\mathrm{a}} = \boldsymbol{\theta}_0$

4. The chi-square estimate is taken as an approximation of ω in that it's an estimate of χ^2 increase (from 0 with the "right" alternative model) only due to the "wrong" constraints

Based on the LM statistic,

- Steps 1-3 are the same as in the LR procedure; then, ω is estimated as the LM statistic
- If only one parameter is considered, ω is simply its modification index

Power used in model evaluation

- Suppose we set $\alpha = 0.05$ and minimum power of 0.7 (see Figure 8.9, p. 347)

Case 1: p-value < 0.05 and estimated power < 0.7 $\rightarrow$ H_0 likely to be false; confidently reject it; significant even with a weak test

Case 2: p-value < 0.05 and estimated power ≥ 0.7 $\rightarrow$ could be ambiguous since suggested rejection may be due to spurious power with very large N

Case 3: p-value ≥ 0.05 and estimated power < 0.7 $\rightarrow$ ambiguous since suggested acceptance may be due to weak test

Case 4: p-value ≥ 0.05 and estimated power ≥ 0.7 $\rightarrow$ H_0 likely to be true; confidently accept it since it's insignificant even with the sufficient power

What to do with too weak test

- Increase α so as to make the test more powerful at the cost of increased type-I error (failure to reject wrong constraints)

- Increase sample size in that chi-square estimate proportional to $N-1$ --- Hoelter's N may be useful, with caution not to make it excessively powerful simply due to too large sample

- Increase # of indicators and/or reliability of given measures so that latent variables become more reliable (i.e., internally consistent), and in consequence estimates are more accurate (so as to increase the chance to reject wrong H_0 when H_a is true) --- note the tradeoff between more reliable measurement by more indicators vs. increase in model df

What to do with excessive power

- “Trivial” deviations may be powerfully detected
- Reduce α so as to make the test more tolerant of Type-II errors (wrong acceptance of false H_0)
- Reducing sample size (e.g., by randomly taking a subsample) is somewhat controversial since it adds more sampling error and consequently yields less accurate estimates by wasting what’s given --- maybe okay only if an optimal N is known so as not to reduce N below the optimal level
- Reducing reliability is more problematic since it amounts to adding measurement errors and yielding “weak” measurement of latent variables

Standardization

- Standardized coefficients useful to compare relative magnitude of parameters --- but blind interpretation without reference to the substantive meaning of "one standard unit change" could be misleading (e.g., 1 SD change of gender)

- To obtain standardized coefficients, multiply unstandardized coefficients by the SD of explanatory variables & divide them by the SD of dependent variables, e.g.,

$$\lambda_{ij}^{(s)} = \lambda_{ij} \left(\frac{\sigma_{jj}}{\sigma_{ii}} \right)^{0.5} \quad \text{or} \quad \mathbf{\Lambda}_x^{(s)} = \mathbf{D}_{xx}^{-0.5} \mathbf{\Lambda}_x \mathbf{D}_{\xi\xi}^{0.5}$$

When $\hat{\sigma}_{ii} \neq s_{ii}$, standardized coefficients may differ by programs

Mean structure

- To include mean structure, all equations need to have an additional term for intercept (for each DV) and all explanatory variables have expectation other than 0, while all error terms are expected 0

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} = \left(\mathbf{I} - \mathbf{B}\right)^{-1}\left(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}\right)$$

$$E\left(\boldsymbol{\eta}\right) = \left(\mathbf{I} - \mathbf{B}\right)^{-1}\left(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\kappa}\right), \quad E\left(\boldsymbol{\xi}\right) = \boldsymbol{\kappa}$$

$$\mathbf{y} = \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad E\left(\mathbf{y}\right) = \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y\left(\mathbf{I} - \mathbf{B}\right)^{-1}\left(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\kappa}\right)$$

$$\mathbf{x} = \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \quad E\left(\mathbf{x}\right) = \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x\boldsymbol{\kappa}$$

Identification for mean-structure parameters

- Input data have $p+q$ more distinctive data df --- observed variables' means

- New parameters in the model: $p+q$ intercepts and $m+n$ means of $\boldsymbol{\eta}$ and ξ; So, we need at least $m+n$ more constraints which fix the shift ambiguity of latent variable distributions --- typically done by setting $\boldsymbol{\kappa}=\mathbf{0}$ and $\boldsymbol{\alpha}=\mathbf{0}$ (and so all LVs have a zero mean) or alternatively one v_x and v_y set to 0 per LV

- With the $m+n$ mean (or intercept) constraints, the mean structure is just identified, and thus not so much of interest at least from the modeling perspective (unless some further constraints are imposed)

- When multiple groups or times (e.g., in panel data) are considered, particularly with some constraints across groups, it's not optimal to arbitrarily choose the metric of LVs (e.g., same metric as one indicator by setting its loading to 1 and intercept to 0) --- an optimal procedure to be discussed in multiple groups analysis (see e-copy of "comparing populations.pdf"; McDonald, R.P., 1999, chapter 15, pp. 325-346 in *Test theory: A unified treatment*)

- Differences in mean structure can be tested by the LR statistic; Or individual differences may be tested by asymptotic variances and covariances of individual estimates --- when requested, AMOS prints critical ratios of differences of all pairs of parameters