General structural model – Part 2: Nonnormality

Conditions for efficient ML & GLS

- F_{ML} is derived with an assumption that all DVs are multivariate normal

- Somewhat relaxed condition (due to Browne) which is satisfied unless the distribution is excessively kurtotic:

$$\mathrm{acov}\left(s_{ij}, s_{gh}\right) = N^{-1}\left(\sigma_{ig}\sigma_{jh} + \sigma_{ig}\sigma_{jh}\right)$$

- Exogenous observed variables don’t have to be multivariate normal so long as all other observed variables are so

- Violation of multinormality or the ACOV condition doesn't cost consistency of the ML and GLS --- but it does make the estimators inefficient and chi-square testing invalid (as well as individual SE estimates) (see Table 9.1, p. 416)

Detecting nonnormality by moments

- The r-th moment around the mean defined as ($r > 1$):

$$\mu_r = E\left(X - \mu_1\right)^r, \qquad m_r = N^{-1}\left(X - \bar{X}\right)^r$$

- When standardized ($\mu_3 \big/ \mu_2^{3/2}, \;\; \mu_4 \big/ \mu_2^2$ for the third and fourth), all moments are mutually <u>uncorrelated</u> --- e.g., larger mean does not imply anything whatsoever about variance

- Multivariate normal distribution has two parameter sets --- mean-vector and covariance matrix, $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- Any higher <u>standardized</u> moments are constants under normality --- e.g., skewness = 0 and kurtosis = 3 (excess kurtosis = 0)

Skewness and kurtosis

$$\text{skewness}: \quad b_1^* = \frac{m_3}{m_2^{3/2}}, \qquad \text{kurtosis}: \quad b_2 = \frac{m_4}{m_2^2}$$

- Skewness --- degree of positive or negative tendency, deviating from normality
  - Positively skewed --- tends more toward positive infinity
  - Negatively skewed --- tends more toward negative infinity

- Kurtosis --- degree of tailedness, deviating from normality
  - Leptokurtic or super-Gaussian $(b_2 > 3)$ --- thicker tail, taller than normality (Fig. 9.3b – A)
  - Platykurtic or sub-Gaussian $(b_2 < 3)$ --- thinner tail, shorter than normality (Fig. 9.3b – C)

Test for nonnormality

- Large sample based z-tests available separately for skewness $(b_1 = 0)$ or kurtosis $(b_2 = 3)$ and simultaniously for both $(b_1 = 0 \ \& \ b_2 = 3)$, for both univariate and multivariate normality (see Tables 9.2 & 9.3)

- Univariate test may be used for identifying a subset of nonnormal variables, one at a time

- Identification of cases deviating from normality (outliers)

  - Deviation from density expected under normality (Q-Q plot)

  - Maharanobis distance (a.k.a., statistical distance) --- useful for identifying extreme cases

- Univariate normality (marginal distributions) is necessary for multinormality

- Most SEM programs provide univariate and multivariate tests for nonnormality; and for outliers (e.g., Maharanobis distance) --- available in AMOS

- Practically, if removing a few outliers reasonably approximates multinormality (or at least, univariate and bivariate normalities), then usual statistical practice can be considered justified; otherwise, some alternative procedure is in need

Corrections

- Transformation of data; e.g., taking logarithm alleviates impact of extremely large values --- results should be accordingly interpreted (e.g., effect of log income instead of income itself)

- Robust statistics for asymptotically valid statistical testing --- not so efficient with smallish samples

- Nonparametric test of overall fit (e.g., bootstrapping) --- known to be erratic sometimes with smallish samples

- Alternative estimator that doesn’t require particular distributional form and, yet, is efficient, given sufficiently large data

Weighted least squares (WLS)

$$F_{\text{WLS}} = \left(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\right)' \mathbf{W}^{-1} \left(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\right), \quad \mathbf{s} : \tfrac{1}{2}(p+q)(p+q+1) \times 1$$

$$\operatorname{acov}\left(s_{ij}, s_{gh}\right) = N^{-1}\left(\sigma_{ijgh} - \sigma_{ij}\sigma_{gh}\right)$$

$$F_{\text{GLS}} = \tfrac{1}{2}\operatorname{tr}\left(\left(\left(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)\mathbf{S}^{-1}\right)^{2}\right) \qquad \mathbf{S} : (p+q) \times (p+q)$$

$$\operatorname{acov}\left(s_{ij}, s_{gh}\right) = N^{-1}\left(\sigma_{ig}\sigma_{jh} + \sigma_{ih}\sigma_{jg}\right), \quad \text{if normal}$$

$$F_{\text{ULS}} = \tfrac{1}{2}\operatorname{tr}\left(\left(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)^{2}\right)$$

- Under normality, F_{WLS} reduces to F_{GLS} and F_{ULS}, respectively with $\mathbf{W} = \mathbf{S}$ and $\mathbf{I}$ (due to Browne)

- Likewise, F_{ML} is also another special case of F_{WLS} under normality, with $\mathbf{W} = \hat{\boldsymbol{\Sigma}}$

- WLS also known as "asymptotically/arbitrary distribution free" (ADF) estimator in that the ACOV holds without needing a particular distributional form, for which

$$\hat{\sigma}_{ijgh} = N^{-1}\sum_{t=1}^{N}\left(X_{it} - \bar{X}_{i}\right)\left(X_{jt} - \bar{X}_{j}\right)\left(X_{gt} - \bar{X}_{g}\right)\left(X_{ht} - \bar{X}_{h}\right)$$

- F_{ADF} uses $\mathbf{W} = \{\mathrm{acov}(s_{ij}, s_{gh})\}_{k\times k}$, which is a square matrix of order $k = (p + q)(p + q + 1)/2$, that needs to be inverted during optimization

- See, e.g., Table 9.4, p. 428

- WLS should not be confused with the WLS for an adjustment for heterogeneous error variances in multiple regression analysis:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

$$\hat{\boldsymbol{\beta}}_{\mathrm{WLS}} = \left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y},$$

$$\mathbf{W} = \mathrm{diag}\left(s_1^2, \ldots, s_N^2\right)$$

WLS pros & cons

- Pros:
  - No distributional form required
  - Efficient estimates (minimum SE)
  - Correct chi-square testing
- Cons:
  - Computationally expensive --- large (and full) $\mathbf{W}$ needs to be iteratively inverted
  - Usual recommendation for minimum sample size: 400~500; properties with small samples unknown
  - Hard to know which performs better (WLS vs. ML or GLS) given not so large sample and/or significant, yet not excessive, nonnormality

Elliptical estimator

- Elliptical distribution has 0 skewness but can be kurtotic by the same degree for <u>all variables</u>:

$$K = \frac{m_{iiii}}{3m_{ii}^2} - 1, \qquad i = 1, \ldots, (p+q)$$

- If the common-kurtosis condition is met, estimates by F_{E} are efficient and results in correct statistical testing --- within this condition, F_{ML} and F_{GLS} are special cases of $K = 0$

- To use F_{E} needs an estimate of K (see Eq. 9.93) --- Mardia's multivariate b_2 or average of univariate b_2's may be used

- Computationally less demanding than WLS