

Second-Order Bias-Corrected AIC for Selecting Structural Equation Models

Kentaro HAYASHI

Department of Psychology, University of Hawaii at Manoa

(E-mail: hayashik@hawaii.edu)

AND

Hirokazu YANAGIHARA

Department of Mathematics, Hiroshima University

(E-mail: yanagi@math.sci.hiroshima-u.ac.jp)

Introduction

We derive a second-order bias correction of Akaike Information Criterion (AIC) in structural equation models (SEM) under the normal assumption when the model is **overspecified**.

Note: “Overspecified” means the candidate models (f 's) include the true model (φ).

Contents:

1. Introduction

(a) Structural Equation Models (SEM)

(b) AIC and CV (Cross-Validation) Criterion

2. General Theory

- (a) True and Candidate Models
- (b) Likelihood and MLE
- (c) Risk, Bias, and Information Criterion
- (d) Estimated Bias

3. Recent and Current Studies

- (a) Notations (Derivatives, Expectation of Moment Matrices, Estimates of Expected Moment Matrices, and Coefficients in Bias-Correction Terms)
- (b) Evaluating Bias of Information Criteria
- (c) Asymptotic Expansion of Expectation of Estimated “Beta” Term
- (d) Bias of AIC (Main Result of Current Study)
- (e) Useful Formulas in Obtaining the Coefficient Terms

Structural Equation Models (SEM)

References: Bollen (1989), Bartholomew and Knott (1999), Skrondal and Rabe-Hesketh (2004), Yuan and Bentler (2007)

- SEM is one of the most frequently used multivariate techniques in social sciences.
- SEM aims to express the covariance structure using relatively small number of parameters. Notation: $\Sigma(\theta)$

- The single most famous SEM is the confirmatory factor analysis (CFA) model:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\varepsilon},$$

where

\mathbf{y} ($p \times 1$): Observed variables,

$\boldsymbol{\mu}$ ($p \times 1$): Population means,

$\boldsymbol{\Lambda}$ ($p \times m$): Factor loadings (Path coefficients),

\mathbf{f} ($m \times 1$): Factors,

$\boldsymbol{\varepsilon}$ ($p \times 1$): Errors.

Note: CFA is a linear model but the factors (\mathbf{f}) are **latent** (NOT observed) variables.

- Typical assumptions: Errors are mutually uncorrelated, and factors and errors are uncorrelated. That is, $\text{Cov}(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = 0$ and $\text{Cov}(\boldsymbol{f}_i, \boldsymbol{\varepsilon}_j) = 0$.
- The covariance structure of variables (\boldsymbol{y}) is expressed as a function of:
 - $\boldsymbol{\Lambda}$ ($p \times m$): Factor loadings (Path coefficients),
 - $\boldsymbol{\Phi}$ ($m \times m$): Factor correlations, and
 - $\boldsymbol{\Psi}$ ($p \times p$): Error variances.
- That is, the covariance structure under CFA is expressed as:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi},$$

where $\boldsymbol{\theta} = (\boldsymbol{\lambda}', \boldsymbol{\phi}', \boldsymbol{\psi}')' = (\theta_1, \dots, \theta_q)'$.

AIC (Akaike Information Criterion)

- When the candidate model is *overspecified*, AIC is the first-order bias-corrected estimator of the risk function based on the expected predictive Kullback-Leibler (KL) discrepancy between the true model and the candidate model. That is,

$$E[\text{AIC}] = R_{\text{KL}} + O(n^{-1}).$$

- AIC tends to choose the model with many parameters as the best model (when the full model has too many parameters).
- *Reason:* AIC tends to underestimate the bias when the candidate model has many parameters (because the bias term of AIC is derived based on the asymptotic theory of $\hat{\theta}$).

- The (negative) property of AIC that the candidate model having too many parameters is chosen as the best model tends to appear in the *overspecified* models.

Cross-Validation (CV)

- Even when the candidate model is misspecified, by correcting the bias of the cross validation (CV) criterion (Stone, 1974), the *second-order bias-corrected* estimators of the risk function have been proposed under the general condition (e.g., Yanagihara, Tonda and Matsumoto, 2006). That is,

$$E[\text{CCV}] = R_{\text{KL}} + O(n^{-2}).$$

- However, the many computational tasks are need for obtaining bias-corrected criteria and these criteria have large variance.

True and Candidate Models (General)

Let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be p -dimensional random observation vectors, where n is the sample size.

- **The true model:** $M_\varphi : \mathbf{y}_1, \dots, \mathbf{y}_n \sim i.i.d. \varphi(\mathbf{y}),$

where $\varphi(\mathbf{y})$ is an *unknown* probability density function.

- **The candidate model:** $M_f : \mathbf{y}_1, \dots, \mathbf{y}_n \sim i.i.d. f(\mathbf{y} | \boldsymbol{\theta}),$

where $\mathcal{F} = \{f(\mathbf{y} | \boldsymbol{\theta}); \boldsymbol{\theta} \subseteq \Theta\}, \boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$.

Candidate Model in SEM

If the candidate model is specified, then

$$NS \sim W_p(N, \Sigma(\boldsymbol{\theta}_0))$$

$$\Rightarrow NS = \mathbf{W}_1 + \cdots + \mathbf{W}_N, \mathbf{W}_1, \dots, \mathbf{W}_N \sim i.i.d. W_p(1, \Sigma(\boldsymbol{\theta}_0)),$$

where $N = n - 1$.

$\mathbf{W}_1, \dots, \mathbf{W}_N$ can be regarded as independent observations. Therefore, the candidate model is:

$$M_f : \mathbf{W}_1, \dots, \mathbf{W}_N \sim i.i.d. W_p(1, \Sigma(\boldsymbol{\theta})).$$

Likelihood and MLE (General)

- **Log-likelihood function:** $L(\boldsymbol{\theta} \mid \mathbf{Y}) = \sum_{i=1}^n \log f(\mathbf{y}_i \mid \boldsymbol{\theta}),$

where $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)'$.

- **Maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$:** $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta} \mid \mathbf{Y})$

- Convergence of MLE in the misspecified model (White, 1982) :

$$\lim_{n \rightarrow \infty} \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0, \quad E_{\mathbf{y}} \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{y} \mid \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \right] = \mathbf{0}$$

where $E_{\mathbf{y}}$ denotes an expectation with respect to \mathbf{y} under the true model $\varphi(\mathbf{y})$.

Likelihood and MLE in SEM

In SEM, the discrepancy function is:

$$F_{\text{KL}}(\boldsymbol{\theta} \mid \boldsymbol{S}) = -\log |\boldsymbol{S}\boldsymbol{\Sigma}^{-1}| + \text{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) - p$$

$$\Rightarrow \frac{1}{N} \sum_{i=1}^N F_{\text{KL}}(\boldsymbol{\theta} \mid \boldsymbol{W}_i).$$

Therefore, the log-likelihood is: $-2L(\boldsymbol{\theta} \mid \boldsymbol{S}) = (N)F_{\text{KL}}(\boldsymbol{\theta} \mid \boldsymbol{S})$,
and the MLE is:

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta} \mid \boldsymbol{S}).$$

Risk Function, Bias, and Information Criterion

- **Risk function based on the expected predictive KL discrepancy:**

$$R_{\text{KL}} = E_y E_u \left[-2L(\hat{\boldsymbol{\theta}} \mid \mathbf{U}) \right],$$

where $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_n)'$ is an $n \times p$ future observation matrix (independent of \mathbf{Y}), and \mathbf{u}_i is distributed according to the same distribution of \mathbf{y}_i ($i = 1, \dots, n$).

- **Bias:** $B = R_{\text{KL}} - E_y \left[-2L(\hat{\boldsymbol{\theta}} \mid \mathbf{Y}) \right]$.

- Information criterion (IC):

$$\text{IC} = -2L(\hat{\boldsymbol{\theta}} \mid \mathbf{Y}) + \hat{B},$$

where \hat{B} is a consistent estimator of B .

Note: The ICs are specified by different terms of \hat{B} .

Estimated Bias in Information Criteria

- AIC: $\hat{B} = 2q$.
- TIC (Takeuchi information criterion; Takeuchi, 1976):

$$\hat{B} = 2\text{tr}\{\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}})\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})^{-1}\}.$$

- CV: $\hat{B} = -2\sum_{i=1}^n \log f(\mathbf{y}_i | \hat{\boldsymbol{\theta}}_{[-i]}) + 2L(\hat{\boldsymbol{\theta}} | \mathbf{Y})$,

where $\hat{\boldsymbol{\theta}}_{[-i]}$ is Jackknife estimator of $\boldsymbol{\theta}$ defined by

$$\hat{\boldsymbol{\theta}}_{[-i]} = \arg \max_{\boldsymbol{\theta}} \left\{ \sum_{j \neq i}^n \log f(\mathbf{y}_j | \boldsymbol{\theta}) \right\}.$$

Notations

A. Derivatives:

1. *First-order (Gradient)*: $g(\mathbf{y} \mid \boldsymbol{\theta}) = -\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{y} \mid \boldsymbol{\theta}),$

2. *Second-order (Hessian)*: $H(\mathbf{y} \mid \boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log f(\mathbf{y} \mid \boldsymbol{\theta}),$

3. *Third-order*: $C(\mathbf{y} \mid \boldsymbol{\theta}) = -\left(\frac{\partial}{\partial \boldsymbol{\theta}'} \otimes \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \log f(\mathbf{y} \mid \boldsymbol{\theta}),$

4. *Fourth-order*: $Q(\mathbf{y} \mid \boldsymbol{\theta}) = -\left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \otimes \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \log f(\mathbf{y} \mid \boldsymbol{\theta}),$

where \otimes is the Kronecker product.

B. Expectation of Moment Matrices

1. **Information:** $I(\boldsymbol{\theta}) = E_y[\mathbf{g}(\mathbf{y} | \boldsymbol{\theta})\mathbf{g}(\mathbf{y} | \boldsymbol{\theta})']$,

2. **Jacobian:** $\mathbf{J}(\boldsymbol{\theta}) = E_y[\mathbf{H}(\mathbf{y} | \boldsymbol{\theta})]$,

3. Expected *third*-order moment matrix: $\mathbf{K}(\boldsymbol{\theta}) = E_y[\mathbf{C}(\mathbf{y} | \boldsymbol{\theta})]$,

4. Expected *fourth*-order moment matrix: $\mathbf{L}(\boldsymbol{\theta}) = E_y[\mathbf{Q}(\mathbf{y} | \boldsymbol{\theta})]$.

C. Estimates of Expected Moment Matrices:

1. Estimated **Information**: $\hat{I}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{g}(\mathbf{y}_i | \boldsymbol{\theta}) \mathbf{g}(\mathbf{y}_i | \boldsymbol{\theta})'$,

2. Estimated **Jacobian**: $\hat{J}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{H}(\mathbf{y}_i | \boldsymbol{\theta})$,

3. Estimated 3rd-order moment: $\hat{K}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{C}(\mathbf{y}_i | \boldsymbol{\theta})$,

4. Estimated 4th-order moment: $\hat{L}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{Q}(\mathbf{y}_i | \boldsymbol{\theta})$.

D(a). Coefficients in Bias-Correction Terms (General Case)

$$\alpha_1 = \text{tr}\{I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}\},$$

$$\alpha_2 = E_y[\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)'J(\boldsymbol{\theta}_0)^{-1}H(\mathbf{y} | \boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)],$$

$$\alpha_3 = E_y[\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)'J(\boldsymbol{\theta}_0)^{-1}K(\boldsymbol{\theta}_0)\{J(\boldsymbol{\theta}_0)^{-1}\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \otimes J(\boldsymbol{\theta}_0)^{-1}\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)\}],$$

$$\alpha_4 = E_y\left[\text{tr}\left\{J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}H(\mathbf{y} | \boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}H(\mathbf{y} | \boldsymbol{\theta}_0)\right\}\right],$$

$$\alpha_5 = E_y\left[\text{tr}\left\{C(\mathbf{y} | \boldsymbol{\theta}_0)(J(\boldsymbol{\theta}_0)^{-1}\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \otimes J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1})\right\}\right],$$

$$\alpha_6 = E_y\left[\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)'J(\boldsymbol{\theta}_0)^{-1}H(\mathbf{y} | \boldsymbol{\theta}_0)\right]J(\boldsymbol{\theta}_0)^{-1}E_y\left[H(\mathbf{y} | \boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)\right],$$

$$\alpha_7 = E_y\left[\text{tr}\left\{K(\boldsymbol{\theta}_0)(J(\boldsymbol{\theta}_0)^{-1}H(\mathbf{y} | \boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \otimes J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1})\right\}\right],$$

$$\alpha_8 = \text{tr}\left\{K(\boldsymbol{\theta}_0)(J(\boldsymbol{\theta}_0)^{-1}K(\boldsymbol{\theta}_0)\text{vec}(J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}) \otimes J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1})\right\},$$

$$\alpha_9 = E_\varphi\left[\mathbf{g}(\mathbf{y}_1 | \boldsymbol{\theta}_0)'J(\boldsymbol{\theta}_0)^{-1}H(\mathbf{y}_2 | \boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}H(\mathbf{y}_1 | \boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}\mathbf{g}(\mathbf{y}_2 | \boldsymbol{\theta}_0)\right],$$

$$\alpha_{10} = E_\varphi\left[\text{tr}\left\{K(\boldsymbol{\theta}_0)(J(\boldsymbol{\theta}_0)^{-1}\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \otimes J(\boldsymbol{\theta}_0)^{-1}H(\mathbf{y} | \boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1})\right\}\right],$$

$$\alpha_{11} = \text{tr}\left\{K(\boldsymbol{\theta}_0)'J(\boldsymbol{\theta}_0)^{-1}K(\boldsymbol{\theta}_0)(J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1} \otimes J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1})\right\},$$

$$\alpha_{12} = \text{tr}\left\{L(\boldsymbol{\theta}_0)(J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1} \otimes J(\boldsymbol{\theta}_0)^{-1}I(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1})\right\}.$$

D(b). Case with Overspecified Models $I(\boldsymbol{\theta}_0) = J(\boldsymbol{\theta}_0)$

$$\alpha_1 = q,$$

$$\alpha_2 = E_{\mathbf{y}}[\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)' \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\mathbf{y} | \boldsymbol{\theta}_0) \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)],$$

$$\alpha_3 = E_{\mathbf{y}}[\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)' \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{K}(\boldsymbol{\theta}_0) \{ \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \otimes \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \}],$$

$$\alpha_4 = E_{\mathbf{y}} \left[\text{tr} \left\{ \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\mathbf{y} | \boldsymbol{\theta}_0) \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\mathbf{y} | \boldsymbol{\theta}_0) \right\} \right],$$

$$\alpha_5 = E_{\mathbf{y}} \left[\text{tr} \left\{ \mathbf{C}(\mathbf{y} | \boldsymbol{\theta}_0) (\mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \otimes \mathbf{I}(\boldsymbol{\theta}_0)^{-1}) \right\} \right],$$

$$\alpha_6 = E_{\mathbf{y}} \left[\mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0)' \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\mathbf{y} | \boldsymbol{\theta}_0) \right] \mathbf{I}(\boldsymbol{\theta}_0)^{-1} E_{\mathbf{y}} \left[\mathbf{H}(\mathbf{y} | \boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0)^{-1} \mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \right],$$

$$\alpha_7 = E_{\mathbf{y}} \left[\text{tr} \left\{ \mathbf{K}(\boldsymbol{\theta}_0) (\mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\mathbf{y} | \boldsymbol{\theta}_0) \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \otimes \mathbf{I}(\boldsymbol{\theta}_0)^{-1}) \right\} \right],$$

$$\alpha_8 = \text{tr} \left\{ \mathbf{K}(\boldsymbol{\theta}_0) (\mathbf{J}(\boldsymbol{\theta}_0)^{-1} \mathbf{K}(\boldsymbol{\theta}_0) \text{vec}(\mathbf{J}(\boldsymbol{\theta}_0)^{-1} \mathbf{I}(\boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0)^{-1}) \otimes \mathbf{I}(\boldsymbol{\theta}_0)^{-1}) \right\},$$

$$\alpha_9 = E_{\mathbf{y}} \left[\mathbf{g}(\mathbf{y}_1 | \boldsymbol{\theta}_0)' \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\mathbf{y}_2 | \boldsymbol{\theta}_0) \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\mathbf{y}_1 | \boldsymbol{\theta}_0) \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{g}(\mathbf{y}_2 | \boldsymbol{\theta}_0) \right],$$

$$\alpha_{10} = E_{\mathbf{y}} \left[\text{tr} \left\{ \mathbf{K}(\boldsymbol{\theta}_0) (\mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{g}(\mathbf{y} | \boldsymbol{\theta}_0) \otimes \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\mathbf{y} | \boldsymbol{\theta}_0) \mathbf{I}(\boldsymbol{\theta}_0)^{-1}) \right\} \right],$$

$$\alpha_{11} = \text{tr} \left\{ \mathbf{K}(\boldsymbol{\theta}_0)' \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{K}(\boldsymbol{\theta}_0) (\mathbf{I}(\boldsymbol{\theta}_0)^{-1} \otimes \mathbf{I}(\boldsymbol{\theta}_0)^{-1}) \right\},$$

$$\alpha_{12} = \text{tr} \left\{ \mathbf{L}(\boldsymbol{\theta}_0) (\mathbf{I}(\boldsymbol{\theta}_0)^{-1} \otimes \mathbf{I}(\boldsymbol{\theta}_0)^{-1}) \right\}.$$

Bias of AIC (Main Result of Current Study)

$$B_{\text{AIC}} = 2(\alpha_1 - q) + \frac{1}{n} \left(-\alpha_1 - 3\alpha_2 + \alpha_3 + 4\alpha_4 + 4\alpha_5 + 4\alpha_6 - 4\alpha_7 \right. \\ \left. + \alpha_8 + 4\alpha_9 - 8\alpha_{10} + 2\alpha_{11} - \alpha_{12} \right) + O(n^{-2}).$$

Useful Formulas in Obtaining the Coefficient Terms

If $\mathbf{W} \sim W_p(1, \Sigma(\boldsymbol{\theta}_0))$, for any symmetric matrices \mathbf{A} , \mathbf{B} , and \mathbf{C} ,

1. $E[\text{tr}(\mathbf{W}\mathbf{A})] = \text{tr}(\mathbf{A})$,

2. $E[\text{tr}(\mathbf{W}\mathbf{A})\text{tr}(\mathbf{W}\mathbf{A})] = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B}) + 2\text{tr}(\mathbf{A}\mathbf{B})$,

3. $E[\text{tr}(\mathbf{W}\mathbf{A})\text{tr}(\mathbf{W}\mathbf{B})\text{tr}(\mathbf{W}\mathbf{C})] = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})\text{tr}(\mathbf{C})$
 $+ 2\{\text{tr}(\mathbf{A})\text{tr}(\mathbf{B}\mathbf{C}) + \text{tr}(\mathbf{B})\text{tr}(\mathbf{A}\mathbf{C}) + \text{tr}(\mathbf{C})\text{tr}(\mathbf{A}\mathbf{B})\} + 8\text{tr}(\mathbf{A}\mathbf{B}\mathbf{C})$.

These expectations imply:

$$4. E[\{\text{tr}(\mathbf{A} - \mathbf{W}\mathbf{A})\}\{\text{tr}(\mathbf{B} - \mathbf{W}\mathbf{B})\}] = 2\text{tr}(\mathbf{A}\mathbf{B}),$$

$$5. E[\{\text{tr}(\mathbf{A} - \mathbf{W}\mathbf{A})\}\{\text{tr}(\mathbf{B} - \mathbf{W}\mathbf{B})\}\{\text{tr}(\mathbf{C} - \mathbf{W}\mathbf{C})\}] = -8\text{tr}(\mathbf{A}\mathbf{B}\mathbf{C}).$$

Note: If \mathbf{B} is not symmetric, then:

$$5(b). \quad E[\{\text{tr}(\mathbf{A} - \mathbf{W}\mathbf{A})\}\{\text{tr}(\mathbf{B} - \mathbf{W}\mathbf{B})\}\{\text{tr}(\mathbf{C} - \mathbf{W}\mathbf{C})\}] \\ = -4\{\text{tr}(\mathbf{A}\mathbf{B}\mathbf{C}) + \text{tr}(\mathbf{A}\mathbf{B}'\mathbf{C})\}.$$

Thank you!