Second-Order Bias-Correction of AIC
for Selecting Structural Equation Models

Kentaro HAYASHI, *Department of Psychology, University of Hawaii at Manoa*
Hirokazu YANAGIHARA, *Department of Mathematics, Hiroshima University*

We derived a second-order bias correction of Akaike Information Criterion (AIC) in structural equation models (SEM) under the normal assumption when the model is overspecified. Structural equation models (SEM) are among the most frequently used in social sciences. The goal of SEM is to express, more exactly, approximate the covariance structure of observed variables using relatively small number of parameters.

The risk function based on the expected Kullback-Leibler discrepancy between the true model and the candidate model is: $R_{\mathrm{KL}} = E_{\boldsymbol{y}} E_{\boldsymbol{u}}\left[-2L(\hat{\boldsymbol{\theta}} \mid \boldsymbol{U})\right]$, where $L(\hat{\boldsymbol{\theta}} \mid \boldsymbol{U})$ is the likelihood of the candidate model, $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)'$ is a matrix of n future observations, and $\boldsymbol{u}_i$ is distributed according to the same distribution as $\boldsymbol{y}_i$. Furthermore, the bias of the risk function is given by: $B = R_{\mathrm{KL}} - E_{\boldsymbol{y}}\left[-2L(\hat{\boldsymbol{\theta}} \mid \boldsymbol{S})\right]$, where $E_{\boldsymbol{y}}[L(\hat{\boldsymbol{\theta}} \mid \boldsymbol{S})]$ is the expected log likelihood. Then, the information criteria can be generically expressed as: $\mathrm{IC} = -2L(\hat{\boldsymbol{\theta}} \mid \boldsymbol{S}) + \hat{B}$, where $\hat{B}$ is a consistent estimator of the bias.

Now, as the notational preparations, we defined the partial derivatives, as follows:

$$\boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}) = -\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\boldsymbol{y} \mid \boldsymbol{\theta}); \boldsymbol{H}(\boldsymbol{y} \mid \boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log f(\boldsymbol{y} \mid \boldsymbol{\theta});$$
$$\boldsymbol{C}(\boldsymbol{y} \mid \boldsymbol{\theta}) = -\left(\frac{\partial}{\partial \boldsymbol{\theta}'} \otimes \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right) \log f(\boldsymbol{y} \mid \boldsymbol{\theta}), \boldsymbol{Q}(\boldsymbol{y} \mid \boldsymbol{\theta}) = -\left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \otimes \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right) \log f(\boldsymbol{y} \mid \boldsymbol{\theta}).$$

Next, we defined the expectation of the moment matrices, as follows:

$$\boldsymbol{I}(\boldsymbol{\theta}) = E_{\boldsymbol{y}}[\boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}) \boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta})']; \boldsymbol{J}(\boldsymbol{\theta}) = E_{\boldsymbol{y}}[\boldsymbol{H}(\boldsymbol{y} \mid \boldsymbol{\theta})]; \boldsymbol{K}(\boldsymbol{\theta}) = E_{\boldsymbol{y}}[\boldsymbol{C}(\boldsymbol{y} \mid \boldsymbol{\theta})];$$
$$\boldsymbol{L}(\boldsymbol{\theta}) = E_{\boldsymbol{y}}[\boldsymbol{Q}(\boldsymbol{y} \mid \boldsymbol{\theta})].$$

Finally, we defined the coefficients in bias-correction terms, as follows:

$$\alpha_1 = \mathrm{tr}\{\boldsymbol{I}(\boldsymbol{\theta}_0) \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\};$$
$$\alpha_2 = E_{\boldsymbol{y}}[\boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}_0)' \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{H}(\boldsymbol{y} \mid \boldsymbol{\theta}_0) \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}_0)];$$
$$\alpha_3 = E_{\boldsymbol{y}}[\boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}_0)' \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{K}(\boldsymbol{\theta}_0) \{\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}_0) \otimes \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}_0)\}];$$
$$\alpha_4 = E_{\boldsymbol{y}}\left[\mathrm{tr}\left\{\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{I}(\boldsymbol{\theta}_0) \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{H}(\boldsymbol{y} \mid \boldsymbol{\theta}_0) \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{H}(\boldsymbol{y} \mid \boldsymbol{\theta}_0)\right\}\right];$$
$$\alpha_5 = E_{\boldsymbol{y}}\left[\mathrm{tr}\{\boldsymbol{C}(\boldsymbol{y} \mid \boldsymbol{\theta}_0)(\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}_0) \otimes \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{I}(\boldsymbol{\theta}_0) \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1})\}\right];$$
$$\alpha_6 = E_{\boldsymbol{y}}\left[\boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}_0)' \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{H}(\boldsymbol{y} \mid \boldsymbol{\theta}_0)\right] \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} E_{\boldsymbol{y}}\left[\boldsymbol{H}(\boldsymbol{y} \mid \boldsymbol{\theta}_0) \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}_0)\right];$$
$$\alpha_7 = E_{\boldsymbol{y}}\left[\mathrm{tr}\left\{\boldsymbol{K}(\boldsymbol{\theta}_0)(\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{H}(\boldsymbol{y} \mid \boldsymbol{\theta}_0) \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{g}(\boldsymbol{y} \mid \boldsymbol{\theta}_0) \otimes \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{I}(\boldsymbol{\theta}_0) \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1})\right\}\right];$$

$$\alpha_8 = \mathrm{tr}\left\{\boldsymbol{K}(\boldsymbol{\theta}_0)(\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{K}(\boldsymbol{\theta}_0)\mathrm{vec}(\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1})\otimes\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1})\right\};$$
$$\alpha_9 = E_\varphi\left[\boldsymbol{g}(\boldsymbol{y}_1\mid\boldsymbol{\theta}_0)'\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{H}(\boldsymbol{y}_2\mid\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{H}(\boldsymbol{y}_1\mid\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{g}(\boldsymbol{y}_2\mid\boldsymbol{\theta}_0)\right];$$
$$\alpha_{10} = E_\varphi\left[\mathrm{tr}\{\boldsymbol{K}(\boldsymbol{\theta}_0)(\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{g}(\boldsymbol{y}\mid\boldsymbol{\theta}_0)\otimes\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{H}(\boldsymbol{y}\mid\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\}\right];$$
$$\alpha_{11} = \mathrm{tr}\{\boldsymbol{K}(\boldsymbol{\theta}_0)'\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{K}(\boldsymbol{\theta}_0)(\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\otimes\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1})\};$$
$$\alpha_{12} = \mathrm{tr}\{\boldsymbol{L}(\boldsymbol{\theta}_0)(\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\otimes\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1})\}.$$

Yanagihara, Yuan, Fujisawa, and Hayashi (2007) derived the bias term of TIC as:

$$B_{\mathrm{TIC}} = R_{\mathrm{KL}} - E_{\boldsymbol{y}}[\mathrm{TIC}] = -\frac{1}{n}(\alpha_1 - 3\alpha_2 + \alpha_3) + O(n^{-2}). \qquad (1)$$

Because AIC and TIC are related with each other as

$$\mathrm{AIC} = \mathrm{TIC} + 2\left\{q - \hat{\beta}(\hat{\boldsymbol{\theta}})\right\} \text{ with } \hat{\beta}(\hat{\boldsymbol{\theta}}) = \mathrm{tr}\{\hat{\boldsymbol{I}}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})^{-1}\}, \qquad (2)$$

it follows that the bias term of AIC can be expressed using the bias term of TIC, as follows:

$$B_{\mathrm{AIC}} = R_{\mathrm{KL}} - E_{\boldsymbol{y}}[\mathrm{AIC}] = B_{\mathrm{TIC}} - 2q + E_{\boldsymbol{y}}[\hat{\beta}(\hat{\boldsymbol{\theta}})] + O(n^{-2}). \qquad (3)$$

Now, because the bias of TIC is given in (1), the remaining task is to find the expression for $E_{\boldsymbol{y}}[\hat{\beta}(\hat{\boldsymbol{\theta}})]$. Thus, we calculated the asymptotic expansion of $E_{\boldsymbol{y}}[\hat{\beta}(\hat{\boldsymbol{\theta}})]$ up to the order n^{-1} in order to obtain the bias of AIC, as follows:

$$\begin{aligned} E_{\boldsymbol{y}}[\hat{\beta}(\hat{\boldsymbol{\theta}})] = \alpha_1 + \frac{1}{n}\Big(&-3\alpha_2 + \alpha_3 + 2\alpha_4 + 2\alpha_5 + 2\alpha_6 - 2\alpha_7 \\ &+\frac{1}{2}\alpha_8 + 2\alpha_9 - 4\alpha_{10} + \alpha_{11} - \frac{1}{2}\alpha_{12}\Big) + O(n^{-2}). \end{aligned} \qquad (4)$$

Therefore, combining (3) and (4), the bias of AIC as the main result of our work is given by:

$$\begin{aligned} B_{\mathrm{AIC}} = 2(\alpha_1 - q) + \frac{1}{n}\big(&-\alpha_1 - 3\alpha_2 + \alpha_3 + 4\alpha_4 + 4\alpha_5 + 4\alpha_6 - 4\alpha_7 \\ &+\alpha_8 + 4\alpha_9 - 8\alpha_{10} + 2\alpha_{11} - \alpha_{12}\big) + O(n^{-2}). \end{aligned} \qquad (5)$$

6. Reference

Yanagihara, H., Yuan, K-H., Fujisawa, H, & Hayashi, K. (2007). A class of model selection criteria based on cross-validation method. *TR 07-01*, *Statistical Research Group*, *Hiroshima University*, Hiroshima.