Unidentified Bi-factor model

January 29, 2021

In this note we show that a very specific bi-factor model is unidentified. This model has 2 specific factors each measured by 3 indicators and 1 general factor measured by all 6 indicators. As usual the factor correlations are fixed to 0 and the factor variances are fixed to 1. The model is given by the following 6 equations

$$Y_1 = \nu_1 + \lambda_1\eta_0 + \beta_1\eta_1 + \varepsilon_1$$

$$Y_2 = \nu_2 + \lambda_2\eta_0 + \beta_2\eta_1 + \varepsilon_2$$

$$Y_3 = \nu_3 + \lambda_3\eta_0 + \beta_3\eta_1 + \varepsilon_3$$

$$Y_4 = \nu_4 + \lambda_4\eta_0 + \beta_4\eta_2 + \varepsilon_4$$

$$Y_5 = \nu_5 + \lambda_5\eta_0 + \beta_5\eta_2 + \varepsilon_5$$

$$Y_6 = \nu_6 + \lambda_6\eta_0 + \beta_6\eta_2 + \varepsilon_6$$

Let

$$Var(\varepsilon_i) = \theta_i.$$

The model has 24 parameters while the unrestricted model has 27. Thus, at first sight the model may appear identified. However, the model is unidentified and the identification problem can be described in two different ways.

The first way to see the non-identification is through the model implied variance covariance matrix V. The variance covariance matrix for the first 3 indicators (which has 6 parameters) can be fitted perfectly with the 6 parameters $\beta_1, \beta_2, \beta_3, \theta_1, \theta_2, \theta_3$. Regardless of what the rest of the model parameters are, those 6 parameters can be used to fit the variance covariance matrix for the first 3 indicators. Similarly, $\beta_4, \beta_5, \beta_6, \theta_4, \theta_5, \theta_6$ can be used to fit the 6 parameters in the variance covariance matrix of indicators Y_4, Y_5 and Y_6.

That means that the 6 parameters λ_i must be identified from the 9 covariance parameters $Cov(Y_i, Y_j) = \lambda_i \lambda_j$, $i = 1, 2, 3$, $j = 4, 5, 6$. Unfortunately that can not be done because for every set of λ_i parameters, the parameters $s\lambda_1$, $s\lambda_2$, $s\lambda_3$, λ_4/s, λ_5/s, λ_6/s yield the same covariances $Cov(Y_i, Y_j)$, where s is any non-zero value. This scale s can not be identified and so the model is not identified.

An alternative way to see the non-identification is as follows. The estimated model necessarily implies the following 4 constraints

$$Cov(Y_1, Y_4) = \frac{Cov(Y_1, Y_6)Cov(Y_3, Y_4)}{Cov(Y_3, Y_6)}$$

$$Cov(Y_1, Y_5) = \frac{Cov(Y_1, Y_6)Cov(Y_3, Y_5)}{Cov(Y_3, Y_6)}$$

$$Cov(Y_2, Y_4) = \frac{Cov(Y_2, Y_6)Cov(Y_3, Y_4)}{Cov(Y_3, Y_6)}$$

$$Cov(Y_2, Y_5) = \frac{Cov(Y_2, Y_6)Cov(Y_3, Y_5)}{Cov(Y_3, Y_6)}.$$

Thus, the estimated model can not have more than 27-4=23 independent parameters. The last 4 sample statistics are dependent quantities. Therefore the bi-factor model with 24 parameters is unidentified. This problem does have a simple solution, however. One of the loadings λ_i can be fixed to the estimated value that Mplus reports in the run that shows the identification problem.

This model non-identification is specific to this particular model. If one of the specific factors has 4 or more indicators the model will be identified.