Model assessment and modification

Contents

1. Introduction

We consider the situation where the researcher has specified a tentative initial model. If the initial model does not fit the given data, the model should be modified and tested again using the same data. Several models may be tested in this process. The goal may be to find a model which not only fits the data well from a statistical point of view, but also has the property that every parameter of the model can be given a substantively meaningful interpretation. The re-specification of each model may be theory-driven or data-driven. Although a model may be tested in each round, the whole approach is *model generating* rather than model testing. In practice, this is a very common situation.

The problem is not just to accept or reject a specified model or to select one out of a set of specified models. Rather, the researcher has specified an initial model that is not assumed to hold exactly in the population and may only be tentative. It’s fit to the data is to be evaluated and assessed in relation to what is known about the substantive area, the quality of the data, and the extent to which various assumptions are satisfied. The evaluation of the model and the assessment of fit is not entirely a statistical matter. If the model is judged not to be good on substantive or statistical grounds, it should be modified within a class of models suitable for the substantive problem. The goal is to find a model within this class of models that not only first the data well statistically, taking all aspects of error into account, but that also has the property of every parameter having a substantively meaningful interpretation.

The output from a structural equations program provides much information useful for model evaluation and assessment of fit. It is helpful to classify this information into the three groups

- Examination of the solution
- Measures of overall fit
- Detailed assessment of fit

and to proceed as follows.

1. Examine the parameter estimates to see if there are any unreasonable values or other anomalies. Parameter estimates should have the right sign and size according to theory or a priori specifications. Examine the squared multiple correlation R^2 for each relationship in the model. The R^2 is a measure of the strength of linear relationship. A small R^2 indicates a weak relationship and suggests that the model is not effective.
2. Examine the measures of overall fit of the model, particularly the chi-square (see Section 2). A number of other measures of overall fit (see Section 3), all of which are functions of the chi-square, may also be used. If any of these quantities indicate a poor fit of the data, proceed with the detailed assessment of fit in the next step.
3. The tools for examining the fit in detail are the residuals and standardized residuals, the modification indices, as well as the expected change (see Section 4). Each of these quantities may be used to locate the source of misspecification and to suggest how the model should be modified to fit the data better.

2. Chi-square

The statistical model and its assumptions imply a covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ for the observed random variables, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_t)$ is a vector of parameters in the statistical model. The testing problem is conceived as testing the model $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. It is assumed that the empirical data is a random sample of size N cases (individuals) on which the observable variables have been observed or measured. From this data a sample covariance matrix **S** is computed, and it is this matrix that is used to fit the model to the data and to test the model.

The model is fitted by minimizing a fit function $F\left[\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right]$ of **S** and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ which is non-negative and zero only if there is a perfect fit, in which case **S** equals $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

Suppose that **S** converges in probability to $\boldsymbol{\Sigma}_0$ as the sample size increases, and let $\boldsymbol{\theta}_0$ be the value of $\boldsymbol{\theta}$ that minimizes $F\left[\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right]$. We say that the model holds if $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. Furthermore, let $\hat{\boldsymbol{\theta}}$ be the value of $\boldsymbol{\theta}$ that minimizes $F\left[\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right]$ for the given sample covariance matrix **S**, and let $n = N - 1$.

To test the model, one may use

$$c = nF\left[\mathbf{S}, \boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)\right]. \quad (1)$$

If the model holds and is identified, c is approximately distributed in large samples as χ^2 with $d = s - t$ degrees of freedom, where $s = k(k+1)/2$ and t is the number of independent parameters estimated. To use this test formally, one chooses a significance level α, draws a random sample of observations on $z_1, z_2, \ldots, z_k$, computes **S**, estimates the model, and rejects the model c exceeds the $(1-\alpha)$ percentile of the χ^2 distribution with d degrees of freedom.

This is a valid test for testing the model against the alternative that the covariance matrix of the observed variables is unconstrained, provided that all assumptions are satisfied, the model holds and the sample size is sufficiently large. In practice it is more useful to regard the chi-square as a *measure* of fit rather than a *test statistic*. In this view, chi-square is a measure of overall fit of the model to the data. It measures the distance (difference, discrepancy, deviance) between the sample covariance (correlation) matrix and the fitted covariance (correlation) matrix. Chi-square is a badness-of-fit measure in the sense that a small chi-square corresponds to good fit and a large chi-square to a bad fit. Zero chi-square corresponds to perfect fit. Chi-square is calculated as $N - 1$ times the minimum value of the fit function, where N is the sample size.

Even if all the assumptions of the chi-square test hold, it may not be realistic to assume that the model holds exactly in the population. In this case, chi-square should be compared with a non-central rather than a central chi-square distribution (see Browne, 1984).

3. Other Goodness-of-Fit Measures

A number of other goodness-of-fit measures have been proposed and studied in the literature. For a summary of these and the rationale behind them, see Bollen (1989). All of these measures are functions of chi-square.

3.1 Goodness-of-fit Indices

Since chi-square is $N - 1$ times the minimum value of the fit function, chi-square tends to be large in large samples if the model does not hold. A number of goodness-of-fit measures have been proposed to eliminate or reduce its dependence on sample size. This is a hopeless task, however, because even though a measure does not depend on sample size explicitly in its calculation, its sampling distribution will depend on N. The goodness-of-fit measures GFI and AGFI of Jöreskog & Sörbom (Tanaka & Huba 1985) do not depend on sample size explicitly and measure how *much* better the model fits as compared to no model at all. For a specific class of models, Maiti & Mukherjee (1990) demonstrate that there is an exact monotonic relationship between GFI and chi-square.

The goodness-of-fit index (GFI) is defined as

$$GFI = 1 - \frac{F\left[\mathbf{S}, \mathbf{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)\right]}{F\left[\mathbf{S}, \mathbf{\Sigma}\left(\mathbf{0}\right)\right]}. \tag{2}$$

The numerator in (2) is the minimum of the fit function after the model has been fitted; the denominator is the fit function before any model has been fitted, or when all parameters are zero.

The goodness-of-fit index adjusted for degrees of freedom, or the adjusted GFI, AGFI, is defined as

$$AGFI = 1 - \frac{k(k+1)}{2d}(1 - GFI), \tag{3}$$

where d is the degrees of freedom of the model. This corresponds to using mean squares instead of total sums of squares in the numerator and denominator of $1 - GFI$. Both of these measures should be between zero and one, although it is theoretically possible for them to become negative. This should not happen, of course, since it means that the model fits worse than no model at all.

3.2 Fit Measures bases on Population Error of Approximation

The use of chi-square as a central χ^2-statistic is based on the assumption tat the model holds exactly in the population. As already pointed out, this may be an unreasonable assumption in most empirical research. A consequence of this assumption is that models which hold approximately in the population will be rejected in large samples. Browne & Cudeck (1993) proposed a number of fit measures which take particular account of the error of approximation in the population and the precision of the fit measure itself. They define an estimate of the *population discrepancy function* as (cf., McDonald, 1989)

$$\hat{F}_0 = Max\{\hat{F} - (d/n), 0\}, \tag{4}$$

where $\hat{F}$ is the minimum value of the fit function, $n = N - 1$ and d is the degree of freedom, and suggest the use of a 90 percent confidence interval

$$(\hat{\lambda}_L / n; \hat{\lambda}_U / n) \tag{5}$$

to assess the error of approximation in the population.

Since $\hat{F}_0$ generally decreases when parameters are added in the model, Browne & Cudeck (1993) suggest using Steiger's (1990) Root Mean Square Error of Approximation (RMSEA)

$$\varepsilon = \sqrt{\hat{F}_0 / d} \tag{6}$$

as a measure of *discrepancy per degree of freedom*. Browne & Cudeck (1993) suggest that a value of 0.05 of ε indicates a close fit and that values up to 0.08 represent reasonable errors of approximation in the population. A 90 percent confidence interval of ε and a test of $\varepsilon < 0.05$ give quite useful information for assessing the degree of approximation in the population. A 90 percent confidence interval for ε is

$$\left(\sqrt{\frac{\hat{\lambda}_L}{nd}}; \sqrt{\frac{\hat{\lambda}_U}{nd}}; \right) \tag{7}$$

and the P-value for test of $\varepsilon < 0.05$ is calculated as

$$P = 1 - G(c \mid 0.0025nd, d) \tag{8}$$

(see Browne & Cudeck, 1993).

3.3 Information Measures of Fit

One disadvantage with chi-square in comparative model fitting is that it always decreases when parameters are added to the model. Therefore there is a tendency to add parameters to the model so as to make chi-square small, thereby capitalizing on chance and ending with a model containing nonsense parameters. A number of measures of fit have been proposed that take parsimony (in the sense of as few parameters as possible) as well as fit into account. These approaches try to deal with this problem by constructing a measure which ideally first decreases as parameters are added and then has a turning point such that it takes its smallest value for the “best” model and then increases when further parameters are added. The AIC and CAIC measures as well as the ECVI belong to this category.

These measures for the estimated model may be compared with the same measures for the “independence” model, *i.e.*, the model in which all observed variables are uncorrelation (if the variables are normally distributed they are independent and their covariance matrix is diagonal; this model has k parameters and k (k - 1)/2 degrees of freedom, where k is the number of observed variables). It will also be useful to compare these measures for the model estimated with those of the saturated model with k (k +1)/2 parameters and zero degrees of freedom. The chi-square for the independence model should also be considered. This provides a test of the hypothesis that the observed variables are uncorrelated. If this hypothesis is not rejected, it is pointless to model the data.

3.4 Other Fit Indices

Another class of fit indices measures how much *better* the model fits as compared to a baseline model, usually the independence model. The first indices of this kind were developed by Tucker & Lewis (1973) and Bentler & Bonett (1980) (NNFI,NFI). Other variations of these have been proposed and discussed by Bollen (1986, 1989) (RFI, IFI) and Bentler (1990) (CFI). These indices are supposed to lie between 0 and 1, but values outside this interval can occur, and, since the independence model almost always has a huge chi-square, one often obtains values very close to 1. James, Mulaik, & Brett (1982, p.155(suggest taking parsimony (degrees of freedom) into account and define a parsimony normed fit index (PNFI), and Mulaik, *et. al.* (1989) suggest a parsimony goodness-of-fit index (PGFI).

For completeness, we give the definition of each of these measures here. Let F be the minimum value of the fit function for the estimated model, let F_i be the minimum value of the fit function for the independence model, and let d and d_i be the corresponding degrees of freedom. Furthermore, let $f = nF/d$, $f_i = nF_i/d_i$, $\tau = \max(nF - d, 0)$, and $\tau_i = \max(nF_i - d_i, 0)$. Then

$$NFI = 1 - F/F_i \tag{9}$$

$$PNFI = (d/d_i)(1 - F/F_i) \tag{10}$$

$$NNFI = \frac{f_i - f}{f_i - 1} \tag{11}$$

$$CFI = 1 - \tau/\tau_i \tag{12}$$

$$IFI = \frac{nF_i - nF}{nF_i - d} \tag{13}$$

$$RFI = \frac{f_i - f}{f_i} \tag{14}$$

$$PGFI = (2d/k(k+1))GFI \tag{15}$$

$$AIC = c + 2t \tag{16}$$

$$CAIC = c + (1 + \ln N)t \tag{17}$$

$$ECVI = (c/n) + 2(t/n)\text{, with 90\% confidence interval } \left(\frac{\lambda_L + \hat{s} + t}{n}; \frac{\lambda_U + \hat{s} + t}{n} \right). \tag{18}$$

Hoelter (1983) proposed a critical N (CN) statistic:

$$CN = \frac{\chi^2_{1-\alpha}}{F} + 1, \tag{19}$$

where $\chi^2_{1-\alpha}$ is the $1-\alpha$ percentile of the chi-square distribution. This is the sample size that would make the obtained chi-square just significant at the significance level α, For a discussion of this statistic, see Bollen & Liang (1988) and Bollen (1989).

The root mean squared residual RMR, is defined as

$$RMR = \left[2\sum_{i=1}^{p+q}\sum_{j=1}^{i}\left(s_{ij} - \hat{\sigma}_{ij} \right)^2 / (p+q)(p+q+1) \right]^{1/2}$$

RMR is a measure of the average of the fitted residuals and can only be interpreted in relation to the sizes of the observed variances and covariances in **S**. This measure works best if all observed variables are standardized.

The root mean squared residual can be used to compare the fit of two different models for the same data. The goodness of fit index can be used for this purpose too but can also be used to compare the fit of models for different data.

It should be emphasized that the measures χ^2, GFI, and RMR are measures of the overall fit of the model to the data and do not express the quality of the model judged by any other internal or external criteria. For example, it can happen that the overall fit of the model is very good but with one or more relationships in the model very poorly determined, as judged by the squared multiple correlations, or vice versa. Furthermore, if any of the overall measures indicate that the model does not fit the data well, it does not tell what is wrong with the model or which part of the model is wrong.

The SRMR is an absolute measure of fit and is defined as the square root of the average of the squared residual correlations. It is a positively biased measure, and that bias is greater for small sample size and for low degrees of freedom studies. The SRMR may be expressed as

$$SRMR = \sqrt{2\sum_{i=1}^{p+q}\sum_{j=1}^{i}\left\{ \frac{s_{ij} - \hat{\sigma}_{ij}}{\sqrt{s_{ii}s_{jj}}} \right\}^2 / (p+1)(P+q+1)}$$

4. Detailed Assessment of Fit

If, on the basis of overall measures of fit or other considerations, it is concluded that the model does not fit sufficiently well, one can examine the fit more closely to determine possible sources of the lack of fit. For this purpose, fitted and standardized residuals and modification indices are useful.

4.1 Fitted and Standardized Residuals

A residual is an observed minus a fitted covariance (variance). A standardized residual is a residual divided by its estimated standard error. There are such residuals for each pair of the observed variables. Fitted residuals depend on the unit of measurement of the observed variables. If the variances of the variables vary considerably from one variable to another, it is rather difficult to know whether a fitted residual should be considered large or small. Standardized residuals, on the other hand, are independent of the units of measurement of the variables. In particular, standardized residuals provide a "statistical" metric for judging the size of a residual.

A large positive residual indicates that the model underestimates the covariance between the variables. In the first case, one should modify the model by adding paths which could account for the covariance between the two variables better. In the second case, one should modify the model by eliminating paths that are associated with the particular covariance.

All the standardized residuals may be examined collectively in two plots: a stemleaf plot and a Q-plot. A good model is characterized by a stemleaf plot in which the residuals are symmetrical around zero, with most in the middle and fewer in the tails. An excess of residuals on the positive or negative side indicates that residuals may be systematically under- or over-estimated in the above sense. In the Q-plot, a good model is characterized by points falling approximately on a 45 degree line. Deviations from this pattern are indicative of specification errors in the model, non-normality in the variables or nonlinear relationships among the variables. In particular, standardized residuals that appear as outliers in the Q-plot are indicative of a specification error in the model.

4.2 Modification Index

A modification index (Sörbom 1989) may be computed for each fixed and constrained parameter in the model. Each such modification index measures how much chi-square is expected to decrease if this particular parameter is set free and the model is reestimated. Thus, the modification index is approximately equal to the difference in chi-square between two models in which one parameter is fixed or constrained in one model and free in the other, all other parameters being estimated in both models. The largest modification index shows the parameter that improves the fit most when set free.

Associated with each modification index, there is an expected parameter change (EPC) (Saris, Satorra, & Sörbom, 1987). This measures how much the parameter is expected to change, in the positive or negative direction, if it is set free. If the units of measurement in observed and/or latent variables are of no particular interest, Kaplan (1989) suggested using a scalefree variant SEPC of EPC.

Modification indices are used in the process of model evaluation and modification in the following way. If chi-square is large relative to the degrees of freedom, one examines the modification indices and relaxes the parameter with the largest modification index *if this parameter can be interpreted substantively*. If it does not make sense to relax the parameter with the largest modification index, one considers the second largest modification index, etc. If the signs of certain parameters are specified a priori, positive or negative, the expected parameter changes associated with the modification indices for these parameters can be used to exclude models with parameters having the wrong sign.

5. Strategy of Analysis

A suitable strategy for data analysis in this situation may be the following.

1. Specify an initial model on the basis of substantive theory, stated hypotheses, or at least some tentative ideas of what a suitable model should be.

2. Estimate the measurement model for each construct separately, then for each pair of constructs, combining them two by two. Then estimate the measurement model for all the constructs without constraining the covariance matrix of the constructs. Finally, estimate the structural equation model for the constructs jointly with the measurement models.
3. For each model estimated in Step 2, evaluate the fit as described in the previous subsections. In particular, pay attention to chi-square, standard errors, *t*-values, standardized residuals, and modification indices. If chi-square is large relative to the degrees of freedom, the model must be modified to fit the data better. For model modification, follow the hints in Section 4. If chi-square is small relative to the degrees of freedom, the model is overfitted and parameters with very large standard errors (very small *t*-values) could possibly be eliminated. If chi-square is in the vicinity of the degrees of freedom, the model may be acceptable, but examine the estimated solution to see if there are any unreasonable values or other anomalies.

   For each model estimated, if this step leads to a modified model, repeat this step on each modified model.

4. Hopefully, the last model estimated in the previous step is one which fits the data of the sample reasonably well and in which all parameters are meaningful and substantively interpretable. However, this does not necessarily mean that it is the "best" model, because its results may have been obtained to some extent by "capitalizing on chance". The model modification process may have generated several "unreasonable" models which should be cross-validated on independent data. If no independent sample is available but the initial sample is large, one may consider splitting the sample into two subsamples and use one (the calibration sample) for exploration as described previously and the other (the validation sample) for cross-validation. The cross-validation is done by computing a validation index (Cudeck & Browne, 1983) for each model. The validation index is a measure of the distance (difference, discrepancy, deviance) between the fitted covariance matrix in the calibration sample and the sample covariance matrix of the validation sample. The model with the smallest validation index is the one which is expected to be most stable in repeated samples. If the smallest validation index occurs for some of the other models that have been fitted, one must make a decision on substantive grounds, which of the two models to retain.

6. Illustration

To illustrate all the goodness-of-fit statistics, we use the following example:

```
                        Goodness-of-Fit Statistics

Degrees of Freedom for (C1)-(C2)                        24
Maximum Likelihood Ratio Chi-Square (C1)                52.626 (P = 0.0006)
Browne's (1984) ADF Chi-Square (C2_NT)                  49.966 (P = 0.0014)

Estimated Non-centrality Parameter (NCP)                28.626
90 Percent Confidence Interval for NCP                  (11.415 ; 53.568)

Minimum Fit Function Value                              0.365
Population Discrepancy Function Value (F0)              0.199
90 Percent Confidence Interval for F0                   (0.0793 ; 0.372)
Root Mean Square Error of Approximation (RMSEA)         0.0910
90 Percent Confidence Interval for RMSEA                (0.0575 ; 0.124)
P-Value for Test of Close Fit (RMSEA < 0.05)            0.0250

Expected Cross-Validation Index (ECVI)                  0.657
90 Percent Confidence Interval for ECVI                 (0.538 ; 0.830)
ECVI for Saturated Model                                0.625
ECVI for Independence Model                             3.571

Chi-Square for Independence Model (36 df)             496.218
```

```
 Normed Fit Index (NFI)                              0.894
 Non-Normed Fit Index (NNFI)                         0.907
 Parsimony Normed Fit Index (PNFI)                   0.596
 Comparative Fit Index (CFI)                         0.938
 Incremental Fit Index (IFI)                         0.939
 Relative Fit Index (RFI)                            0.841

 Critical N (CN)                                   118.606

  Root Mean Square Residual (RMR)                     0.0755
 Standardized RMR                                    0.0755
 Goodness of Fit Index (GFI)                         0.928
 Adjusted Goodness of Fit Index (AGFI)               0.866
 Parsimony Goodness of Fit Index (PGFI)              0.495
```

The chi-square test of exact fit would reject the model, since the *P*-value is very small. Following the guidelines of Browne & Cudeck (1993), it is seen that the point estimate of RMSEA is 0.091 and the 90 % confidence interval is from 0.0575 to 0.124. Since the lower bound is above the recommended value of 0.05, it is concluded that the degree of approximation in the population is too large. So the model is rejected.

Using the modification indices also given in the output

```
        The Modification Indices Suggest to Add the
  Path to          from              Decrease in Chi-Square  New Estimate
         ADDITION          Visual            10.5                 -0.37
         COUNTDOT          Verbal            10.1                 -0.28
         S-C CAPS          Visual            24.7                  0.57
         S-C CAPS          Verbal            10.0                  0.26

 The Modification Indices Suggest to Add an Error Covariance
  Between           and            Decrease in Chi-Square    New Estimate
         COUNTDOT         ADDITION           25.1                  0.62
         S-C CAPS         VIS PERC            9.1                  0.18
         S-C CAPS         COUNTDOT            8.3                 -0.38
```

we note that the largest modification index is 24.7 for the path from Visual to S-C CAPS. This indicates that we can expect a large decrease in chi-square if we include this path in our model. Therefore, *if we can interpret this path substantively*, we can modify the model by adding this path and running the modified model.

The fact that the model is misspecified can also be seen from the standardized residuals in the output file:

```
                 Standardized Residuals

            VIS PERC      CUBES   LOZENGES   PAR COMP   SEN COMP   WORDMEAN
            --------   --------   --------   --------   --------   --------
VIS PERC       - -
    CUBES    -0.793       - -
 LOZENGES    -1.353      2.087       - -
 PAR COMP     0.438     -0.139      0.020       - -
 SEN COMP     0.020     -1.294     -0.565      0.416       - -
 WORDMEAN     0.529     -0.609      0.950     -0.356     -0.054       - -
 ADDITION    -1.946     -1.762     -3.117      0.375      1.225     -0.056
COUNTDOT      0.884     -1.102     -1.141     -3.013     -0.655     -2.084
S-C CAPS      4.650      0.958      2.294      2.241      2.904      1.776

         Standardized Residuals
```

```
            ADDITION   COUNTDOT   S-C CAPS
            --------   --------   --------
ADDITION       - -
 COUNTDOT      5.007       - -
 S-C CAPS     -2.007     -2.886       - -
```

This shows a large standardized residual of 4.650 between VIS PERC and S-C CAPS, indicating that these two variables correlate more than the model accounts for. Although this shows where the lack of fit is, it does not tell how the model should be modified to fit the data better. From this point of view, modification indices are often more useful than standardized residual for detecting specification errors in the model. When the model is modified, the following fit statistics are obtained:

```
                              Goodness-of-Fit Statistics

 Degrees of Freedom for (C1)-(C2)                        23
 Maximum Likelihood Ratio Chi-Square (C1)                28.862 (P = 0.1849)
 Browne's (1984) ADF Chi-Square (C2_NT)                  28.674 (P = 0.1914)

 Estimated Non-centrality Parameter (NCP)                5.862
 90 Percent Confidence Interval for NCP                  (0.0 ; 23.758)

 Minimum Fit Function Value                              0.200
 Population Discrepancy Function Value (F0)              0.0407
 90 Percent Confidence Interval for F0                   (0.0 ; 0.165)
 Root Mean Square Error of Approximation (RMSEA)         0.0421
 90 Percent Confidence Interval for RMSEA                (0.0 ; 0.0847)
 P-Value for Test of Close Fit (RMSEA < 0.05)            0.574

 Expected Cross-Validation Index (ECVI)                  0.506
 90 Percent Confidence Interval for ECVI                 (0.465 ; 0.630)
 ECVI for Saturated Model                                0.625
 ECVI for Independence Model                             3.571

 Chi-Square for Independence Model (36 df)             496.218

 Normed Fit Index (NFI)                                  0.942
 Non-Normed Fit Index (NNFI)                             0.980
 Parsimony Normed Fit Index (PNFI)                       0.602
 Comparative Fit Index (CFI)                             0.987
 Incremental Fit Index (IFI)                             0.988
 Relative Fit Index (RFI)                                0.909

 Critical N (CN)                                       208.748


 Root Mean Square Residual (RMR)                         0.0452
 Standardized RMR                                        0.0452
 Goodness of Fit Index (GFI)                             0.958
 Adjusted Goodness of Fit Index (AGFI)                   0.917
 Parsimony Goodness of Fit Index (PGFI)                  0.489
```

Here the point estimate of RMSEA is below 0.05 and the upper confidence limit is only slightly above the value of 0.08 suggested by Browne & Cudeck (1993). The *P*-value for test of close fit is 0.574. Another indication that the model fits well is that the ECVI for the model (0.506) is less than the ECVI for the saturated model (0.657). In fact, the confidence interval for ECVI is from 0.465 to 0.630. We conclude that the model fits well and represents a reasonably close approximation in the population.

7. References

Akaike H. (1974). A new look at Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Akaike H. (1987). Factor Analysis and AIC. *Psychometrika*, **52**, 317-332.

Bentler, P.M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, **107**, 238-246.

Bentler, P.M. & Bonett, D.G. (1980). Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychological Bulletin*, **88**, 588-606.

Bollen, K.A. (1986). Sample size and Bentler & Bonett's nonnormed fit index. *Psychometrika*, **51**, 375-377.

Bollen, K.A. (1989). *Structural Equation with Latent Variables*. New York: Wiley.

Bollen, K.A. & Liang, J. (1988). Some properties of Hoelter's CN. *Sociological Methods and Research*, **16**, 492-503.

Browne, M.W. (1984). Asymptotically Distribution-free Methods in the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62-83.

Browne, M.W. & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In K.A. Bollen & J.S. Long (Eds.), *Testing Structural Equation Models*, pp. 136-162. Beverley Hills, CA: Sage.

Cudeck, R. & Browne, M.W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, **18**, 147-157.

James, L.R., Mulaik, S.A. & Brett, J.M. (1982). *Causal analysis: Assumptions, model, and data*. Beverly Hills: Sage.

Kaplan, D. (1989). Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research*, **24**, 285-305.

Hoelter, J.W. (1983). The analysis of covariance structures: goodness-of-fit indices. *Sociological Methods and Research*, **11**, 325-344.

Maiti, S.S. & Mukherjee, B.N. (1990). A note on distributional properties of the Jöreskog-Sörbom fit indices. *Psychometrika*, **55**, 721-726.

McDonald. R.P. (1989). An index of goodness-of-fit based on non-centrality. *Journal of Classification*, **6**, 97-103.

Mulaik, S., James, L., Van Alstine, J., Bennett, N, Lind, S & Stilwell, C. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, **105**, 430-445.

Saris, W.E., Satorra, A. & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. Clogg (Ed.): *Sociological Methodology*. San Francisco: Jossey Bass.

Sörbom, D. (1989). Model modification. *Psychometrika*, **54**, 371-384.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, **25**, 173-180.

Tanaka. J.S. & Huba, G.J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, **42**, 233-239.

Tucker, L.R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, **38**, 1-10.