STRUCTURAL EQUATION MODELING, *11*(1), 45–50

A Cautionary Note on the Use of Information Fit Indexes in Covariance Structure Modeling With Means

Jelte M. Wicherts and Conor V. Dolan
Department of Psychology, Psychological Methods
University of Amsterdam

Information fit indexes such as Akaike Information Criterion, Consistent Akaike Information Criterion, Bayesian Information Criterion, and the expected cross validation index can be valuable in assessing the relative fit of structural equation models that differ regarding restrictiveness. In cases in which models without mean restrictions (i.e., saturated mean structure) are compared to models with restricted (i.e., modeled) means, one should take account of the presence of means, even if the model is saturated with respect to the means. The failure to do this can result in an incorrect rank order of models in terms of the information fit indexes. We demonstrate this point by an analysis of measurement invariance in a multigroup confirmatory factor model.

Often in structural equation modeling, a sequence of increasingly restrictive models is fitted. When both means and covariances are modeled, the situation may arise in which one first fits a series of models to the observed covariance matrix, and one subsequently adds the model for the means. Such a stepwise approach has the advantage that it provides information concerning the drop in fit when structured means are added. This is especially important when the means and the covariance structure are modeled with a common subset of parameters (i.e., when strong hypotheses are tested concerning the common causation of individual and mean differences; e.g., Mandys, Dolan, & Molenaar, 1994; Meredith, 1993). The aim of this article is to point out that in the calculation of information criteria and the expected cross validation index (ECVI) in this context one should take into account the presence of means, even if the model is saturated with respect to the means. The failure to do this can result in an incorrect rank order of models by Akaike Information Criterion (AIC; Akaike, 1974), Bayesian Information Crite-

Requests for reprints should be sent to Jelte M. Wicherts, Department of Psychology, Psychological Methods, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands, E-mail: jwicherts@fmg.uva.nl

rion (BIC; Schwarz, 1978), Consistent Akaike Information Criterion (CAIC; Bozdogan, 1987), and ECVI (Browne & Cudeck, 1989, 1993). Specifically the rank order is incorrect when going from a model in which the model for the means is saturated to a model in which the means are constrained. We identify this problem later and demonstrate it in an illustrative analysis.

ASSESSMENT OF RELATIVE FIT USING AIC, BIC, CAIC, AND ECVI

In assessing the fit of structural equation models, it is advisable to consider several fit measures, in addition to the chi-square index (Bollen & Long, 1993). Information criteria such as AIC, CAIC, and BIC form a useful class of indexes, as they penalize for the number of parameters, and thus take into consideration the parsimony of models. Although the information statistics have rather different origins, varying from the concept of entropy (AIC) to Bayesian statistics (BIC), they all have a similar structure (see Table 1) in that they involve the same information. Lower information index values indicate better fit. We note that the ECVI (Browne & Cudeck, 1989, 1993) is linearly related to the AIC, and thus yields the same rank order of competing models as the AIC.

Information statistics are valuable in analyses where models without restrictions on the mean[1] are compared to models with such restrictions. However when means are unrestricted, one may be inclined to discard the means. Clearly, means need not actually be included in a model in which the means are not structured. Moreover, in certain cases (exploratory factor analyses), it is difficult to actually include the means. However, comparing models that do restrict means to models that do not, these implicit mean parameters have to be considered in the computation of the information indexes. If these parameters are overlooked, the information indexes are underestimated. This, in turn, may result in the unjustified rejection of restrictions on the means. The underestimation caused by ignoring the parameters for the means differs for each information criterion and depends on the number of manifest variables and the number of cases. Table 1 contains expressions for this underestimation for each information criterion. We illustrate our point by testing for factorial invariance in two groups of children.

ILLUSTRATION: FACTORIAL INVARIANCE

The psychometric theory concerning the definition and meaning of measurement invariance within the context of the common factor model (i.e., factorial invariance)

[1]That is, a saturated mean structure in which a parameter is estimated for each observed mean.

TABLE 1
Fit Indexes

Fit Index	*Formula*	*Underestimation Due to Ignoring Saturated Means*
AIC	$= \chi^2 + 2t$	$2*p$
CAIC	$= \chi^2 + (1 + \ln N)t$	$(1 + \ln N)*p$
BIC	$= \chi^2 + (\ln N)t$	$(\ln N)p$
ECVI	$= (\chi^2/n) + 2(t/n)$	$2*(p/n)$

Note. t = number of parameters; p = number of manifest means; N = number of cases; n = N-number of groups; AIC = Akaike Information Criterion; CAIC = Consistent Akaike Information Criterion; BIC = Bayesian Information Criterion; ECVI = expected cross validation index.

is well developed (Meredith, 1993). This theory gives rise to multigroup confirmatory factor models in which covariance and mean structures are restricted over groups. Here we compare two groups. Let μ_i and Σ_i denote the implied mean vector and covariance matrix in group i. These are modeled as follows:

$$\mu_i = \nu_i + \Lambda_i \alpha_i \tag{1}$$

$$\Sigma_i = \Lambda_i \Psi_i \Lambda_i' + \Theta_i \tag{2}$$

where the $(p \times q)$ matrix Λ_i contains factor loadings, and the p-dimensional vector ν_i contains measurement intercepts. The $(p \times p)$ diagonal matrix Θ_i contains unique/error variances, and Ψ_i is the $(q \times q)$ covariance matrix of the q common factors. Finally, α_i is a q-dimensional vector of factor means. For reasons of identification (see Sörbom, 1974), this vector is fixed to zero in an arbitrary group so that latent differences in means are modeled. Factorial invariance can be tested by fitting a series of increasingly restricted models. These are presented in Table 2.

In addition to an exploratory factor analysis (EFA), we fit three models without mean restrictions, namely configural invariance (equal pattern of factor loadings), metric invariance (equal factor loadings; Horn, McArdle, & Mason, 1983), and a model with group-invariant error/unique variances. Furthermore, we fit two models with structured means, denoted strong factorial invariance and strict factorial invariance (Meredith, 1993). Meredith showed that, within the factor model, strict factorial invariance is required to demonstrate measurement invariance (i.e., unbiasedness) with respect to groups. To illustrate our point, we fit these models and calculate the indexes with and without taking the means into account.

The models are fitted on a subset of data published in Naglieri and Jensen (1987), which comprise the Kaufman Assessment Battery for Children (K–ABC; Kaufman & Kaufman, 1983) and Wechsler Intelligence Scale for Children–Revised (WISC–R; Wechsler, 1974) scores of 86 Black and 86 White children. We first car-

TABLE 2
Summary of Models in Case of Two Groups 1 and 2

No.	*Description*	$\Sigma_1 =$	$\Sigma_2 =$	$\mu_1 =$	$\mu_2 =$
0	Exploratory	$\Lambda^*_1\Lambda^{*t}_1 + \Theta_1$	$\Lambda^*_2\Lambda^{*t}_2 + \Theta_2$	ν_1	ν_2
1	Configural invariance	$\Lambda_1\Psi_1\Lambda_1^t + \Theta_1$	$\Lambda_2\Psi_2\Lambda_2^t + \Theta_2$	ν_1	ν_2
2	Metric invariance	$\Lambda\Psi_1\Lambda^t + \Theta_1$	$\Lambda\Psi_2\Lambda^t + \Theta_2$	ν_1	ν_2
3	Equal error/unique variances	$\Lambda\Psi_1\Lambda^t + \Theta$	$\Lambda\Psi_2\Lambda^t + \Theta$	ν_1	ν_2
4a	Strict factorial invariance	$\Lambda\Psi_1\Lambda^t + \Theta$	$\Lambda\Psi_2\Lambda^t + \Theta$	ν	$\nu + \Lambda\alpha_2$
4b	Strong factorial invariance	$\Lambda\Psi_1\Lambda^t + \Theta_1$	$\Lambda\Psi_2\Lambda^t + \Theta_2$	ν	$\nu + \Lambda\alpha_2$

Note. Λ^*_1 denotes that all elements are estimated. Except for Step 4b (nested under 2), each model is nested under the previous one.

ried out an EFA on 16 selected subscales (for similar analyses of the complete dataset, see Dolan & Hamaker, 2001). This resulted in a simple structure with three common factors relating to verbal abilities (V), spatial abilities (S), and memory (M). The scales are as follows: Information (loading on the factor V), Similarities (V), Vocabulary (V), Comprehension (V), Picture Completion (S), Picture Arrangement (S), Block Design (S), Object Assembly (S), and Digit Span (M) from the WISC–R; and Faces and Places (V), Riddles (V), Reading/Understanding (V), Triangles (S), Hand Movement (M), Number Recall (M), and Word Order (M) from the K–ABC. In subsequent confirmatory analyses, we use this simple structure. We fix one factor loading per factor at 1 for scaling purposes. Furthermore, we assume multivariate normality and estimate parameters by maximum likelihood.

The fit indexes of the models are presented in Table 3. For comparison, we also report the chi-squares and root mean square error of approximation (RMSEA) fit indexes, which are unaffected by the presence or absence of means in saturated mean models (i.e., models 0–3). We first consider the chi-square indexes. Given the nesting of the models, we employ chi-square differences as a significance test for each restriction (Jöreskog, 1971). This would lead us to conclude that the equality of unique/error variances over groups is not tenable, but that the other between-group restrictions do not lead to a significant decrease ($p < .05$) in chi-square. Based on the chi-square, we therefore conclude that strong factorial invariance holds. Note that the RMSEA does not really help in selecting models. Given the rule of thumb that RMSEA less than .05 represents a reasonable approximation (Browne & Cudeck, 1993), all models are judged to be acceptable. In view of the equivocality of RMSEA, and given the recommendation to consider a variety of indexes (Bollen & Long, 1993), we now turn to the information criteria.

Here we first look at the case in which means are not incorporated in the model (i.e., models 0–3). Based on both the ECVI and the AIC, we would conclude the equality over groups of error/unique variances is not tenable and, more important, that intercepts cannot be equated across groups. The latter also applies to BIC and CAIC, although these two indexes indicate that error/unique variances are invari-

TABLE 3
Fit Indexes of Models With or Without Means

					Means Excluded in 0–3				*Means Included in 0–3*			
Model	*df*	χ^2	*p*	*RMSEA*	*ECVI*	*AIC*	*CAIC*	*BIC*	*ECVI*	*AIC*	*CAIC*	*BIC*
0	150	177.5	.062	0.032	2.39	407	913	791	2.77	471	1,110	956
1	202	233.4	.064	0.020	2.05	349	639	569	2.43	413	836	734
2	215	250.5	.049	0.028	2.02	344	580	523	2.40	408	777	688
3	231	294.1	.003	0.044	2.06	350	520	479	2.44	414	717	644
4a	244	311.3	.002	0.041	2.35	399	648	588	2.35	399	648	588
4b	228	267.4	.038	0.026	2.31	393	708	632	2.31	393	708	632

Note. The chi-square reported here is the minimum fit chi-square, whereas the slightly different Normal Theory Weighted Least Squares chi-square is used here (like it is in LISREL) for computation of the information indexes. RMSEA = root mean square error of approximation; ECVI = expected cross-validation index; AIC = Akaike Information Criterion; CAIC = Consistent Akaike Information Criterion; BIC = Bayesian Information Criterion.

ant across groups. Therefore, when the saturated mean structure is ignored, ECVI, AIC, CAIC, and BIC lead to the incorrect conclusion that both strong and strict factorial invariance should be rejected. Only when the parameters for the means are taken into account (although they are unconstrained) do we draw the correct conclusion. Here, strong factorial invariance does hold, whereas strict factorial invariance would be rejected (e.g., compare the BIC and CAIC in Model 4a and Model 4b; see Table 3).

CONCLUSION

The use of information criteria such as AIC, BIC, CAIC, and the ECVI is valuable in the comparison of structural equation models that differ with respect to restrictiveness. However, when mean structure is analyzed in addition to the covariance structure this mean structure should be incorporated in the models at all stages of model fitting, even when the mean structure is saturated (unrestricted). Failure to do so may result in an incorrect rank order of models and incorrect conclusions. Happily, the correct value of the criteria can be obtained by including the means in the input and model specification.[2] In situations where this may not be possible (e.g., EFA), the correct value can readily be calculated by hand (see Table 1). Although we have focused on factorial invariance in our illustration, this conclusion applies to other models including structured means such as the latent growth curve model or quasisimplex models with structured means (e.g., Mandys et al., 1994). Finally, we note other fit indexes (e.g., the various comparative fit indexes, such as

[2]In LISREL, the ty vector can be used to this end.

the Nonnormed Fit Index) and related information (standardized residuals, modification indexes) are invariant whether or not saturated means are included in the model.

ACKNOWLEDGMENT

The preparation of this article was supported by a grant from the Netherlands Organization for Scientific Research (NWO).

REFERENCES

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control, AC–19,* 716–723.

Bollen, K. A., & Long, J. S. (1993). Introduction. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 1–9). Newbury Park, CA: Sage.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52,* 345–370.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24,* 445–455.

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). Newbury Park, CA: Sage.

Dolan, C. V., & Hamaker, E. L. (2001). Investigating Black–White differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC–R and K–ABC and a critique of the method of correlated vectors. In F. Columbus (Ed.), *Advances in psychology research* (Vol. 6, pp. 31–59). Huntington, NY: Nova Science.

Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist, 1,* 179–188.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409–426.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children.* Circle Pines, MN: American Guidance Service.

Mandys, F., Dolan, C. V., & Molenaar, P. C. M. (1994). Two aspects of the simplex model: Goodness of fit to linear growth curve structures and the analysis of mean trends. *Journal of Educational and Behavioral Statistics, 19,* 201–215.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543.

Naglieri, J. A., & Jensen, A. R. (1987). Comparison and Black–White differences on the WISC–R and the K–ABC: Spearman's hypothesis. *Intelligence, 11,* 21–43.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 229–239.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children–Revised.* New York: The Psychological Corporation.