ELSEVIER Computational Statistics & Data Analysis 41 (2002) 271–287

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Model modification in structural equation modeling by imposing constraints ☆

Chih-Ping Chou[a,*], P.M. Bentler[b]

[a] *Department of Preventive Medicine, University of Southern California, 100 S. Fremont, Unit 8, 91803 Alhambra, CA, USA*

[b] *Department of Psychology, University of Southern California, USA*

Received 1 September 2001; accepted 1 February 2002

Abstract

Two approaches for model modification, or specification search, in structural equation modeling are: releasing constraints (e.g., adding free parameters) and imposing constraints (e.g., deleting free parameters). These two approaches can be considered as "forward search" and "backward search", respectively. The forward search approach, which starts with a more limited model to find a more general model by adding free parameters, is very frequently utilized in practice. However, the success rate of the forward approach in obtaining the correct model is usually low. The forward search also has been found to depend very highly on the initial model selected in the modification process. The farther the initial model differs from the true model, the less likely is the proper modification to occur. This study illustrates the backward search approach, which starts with a more general model and improves the model by imposing constraints, using the z and Wald (W) tests. Assuming a correct measurement model, this study focuses on the model modification of structural relationships among factors. The results showed that with correct specification of the causal order of latent factors the correct model may be located with higher than 60% accuracy, using the incremental univariate Wald test, or the z test. With more cautious, or restricted, search, the success rates are even higher.

Keywords: Structural equation model; Model modification; Backward search

☆ Supported in part by USPHS grants DA00215 and DA01070 from National Institute on Drug Abuse and CA84735 from the National Cancer Institute.

* Corresponding author.

E-mail address: cchou@usc.edu (C.-P. Chou).

PII: S0167-9473(02)00097-X

1. Introduction

Structural equation models (SEM) are usually used in model evaluation situations, such as strictly confirmatory, alternative models, and model generating. Among these situations, model generating in which only a tentative model has been specified as an original model is by far the most common situation (Jöreskog, 1993). In the event that the original model is not appropriate, the model needs to be modified and tested. Two basic approaches for model modification in SEM involve reducing or increasing constraints on the model being tested (Chou and Bentler, 1990). The procedures of reducing and increasing constraints parallel forward and backward stepwise procedures, respectively, in regression analysis. Reducing constraints, or adding free parameters, has been a more common practice in model improvement, especially when the original model does not fit the data. This approach starts with a more limited model and proceeds to a more general model. It represents a type of "forward searching". A problem in releasing constraints to improve a model is that these model modifications or specification searches (Jöreskog and Sörbom, 1988; Leamer, 1978; MacCallum, 1986) for a correct model are not very likely to be successful (MacCallum, 1986). MacCallum and colleagues (MacCallum, 1986; MacCallum et al., 1992; Silvia and MacCallum, 1988) proposed several important suggestions that can lead to more successful specification searches in SEM, such as carefully formulating the initial model, using a large sample, adopting restricted or theoretically guided searches, continuing searches even after a model with a nonsignificant (χ^2) has been obtained, and focusing only on a pre-specified set of fixed parameters that are considered substantively meaningful.

This study describes an evaluation of the alternative approach to model modification for SEM through imposing constraints (Lee, 1984). The process starts with a more general, or saturated, model and aims to find a more restricted model. This procedure of model selection, which is similar to the backward stepwise procedure in regression analysis, is considered as "backward search" approach. Although this approach appears not to have been studied empirically, theoretically, it should be more successful than forward searching since the statistics based on backward search will not be correlated with the fit of the initial more general model (Steiger et al., 1985). It is also likely that starting with an adequate but overparameterized model is liable to lead to more correct model modifications than starting with an inadequate and underparameterized model.

This study intends to evaluate the performance of backward search in the model modification process of SEM against that of the forward search. MacCallum's study on forward specification (1986) represents one of the earliest efforts to investigate the issue of model modification in SEM. His examples will be used for comparisons in this study. To maintain the comparability, assumptions made in his study are also kept in this current study. An important concept, restricted search that assumes that researchers have sufficient knowledge to guide valid modifications, proposed by MacCallum (1986) is also adopted in this study. The definition of restricted search is that "no parameter defined as zero in the true model may be freed, and no nonzero parameter in the true model may be fixed" (MacCallum, 1986, p. 112). Although this definition is too restrictive and too vague in practice since the "true model" is not known empirically

(Gleman et al., 1996), the concept of restricted search, which utilizes as much a priori information as possible during the process of model modification, is important for a successful search. The restricted search also helps to limit the number of possible models that might be considered. More importantly, the restricted search usually requires theoretical justification. Our use of restricted search also requires strong theoretical assumptions will be explained below for the initial models developed in our study. The basic criterion to evaluate the performance of forward and backward searches will be the percentage of locating the correct model.

2. Method

Monte Carlo studies were carried out to examine the extent to which the proposed backward approach can lead to successful specification searches. To compare this backward searching approach through imposing constraints with the more commonly adopted forward searching approach through reducing constraints, the two true models considered in MacCallum's (1986) study were used for data simulation. To maintain the comparability and theoretical meaningfulness, the classification of a latent factor as an independent or a dependent construct was retained. In addition, the recursiveness of a model was also assumed; i.e., only one-way paths were allowed between pairs of dependent latent factors.

2.1. The Wald test and the z test

Two methods for simplifying a model were utilized. These are called the Wald test and z test, respectively. The Wald test (or W test) is a classical test and has become available in structural equation modeling programs (Bentler, 1995). The z test is the standard test based on the value of the parameter estimate divided by standard error to evaluate the hypothesis that the parameter is zero, that is usually printed out in structural modeling programs.

The W Test: The Wald test with maximum likelihood estimation in EQS (Bentler, 1995) was used to reduce free parameters from saturated models in this study. In the multivariate Wald test, EQS provides a stepwise process that enters one parameter at a time into the test procedure (Chou and Bentler, 1996). There are two χ^2 values computed with the inclusion of a parameter at each step. The first χ^2, the overall or multivariate χ^2, is used to test the null hypothesis that the set of parameters included up to the current step is equal to zero (i.e., $\Theta_W = 0$). The z test used in structural equation modeling to test if one parameter is significantly different from zero, $\theta_i = 0$, can be considered as a special case of this W test; the χ^2 of a univariate Wald test is the square of the normal z value. The second χ^2 value, the χ^2 increment, provides information in regard to the contribution of a specific parameter at this step to the overall χ^2. The χ^2 increment is a byproduct of the multivariate Wald test that evaluates $(\theta_i = 0|\Theta_W)$, or $\theta_i^+ = 0$. The χ^2 increment test is, therefore, a conditional test on a specific parameter. It evaluates if $\theta_i = 0$ given that a set of parameters Θ_w has already been included in the testing procedure. Considering a model with q free parameters, Θ_W is an empty set at the first testing of $\theta_i = 0$. It will contain one parameter when testing the second $\theta_i = 0$. The χ^2 increments can continue, in principle, until the very last parameter in the model, when Θ_w will have $(q-1)$ elements, or parameters. In this study, however, the 0.05 probability value (p-value) is used for all statistical decisions. The W test procedure begins by finding the parameter whose estimated z test is smallest, and, accordingly, has the largest p-value. Given that this parameter is in the W test, the procedure searches for the single parameter to be added that makes the smallest increment to the W test. With two parameters included in the W test, the procedure continues to drop free parameters (include them in the W test) only as long as the probability associated with the univariate increment in the W test is > 0.05.

Although both multivariate χ^2 and univariate χ^2 increment values are important, the latter is more useful in deciding whether to drop parameters from the model. The overall χ^2 shows the importance, or rather the unimportance, of a set of parameters in the model. The χ^2 increment, which takes the difference of the overall χ^2 values between the current step and the previous one, reveals the impact of removing a parameter after the parameters in the previous steps have been extracted from the model. In standard hypothesis-testing situations, comparing two a priori models, under the null hypothesis the W test behaves in large samples like the χ^2 difference test (Satorra, 1989). This known theory is usually not relevant where the W test is used in a post-hoc fashion such as in this study. However, the W test conducted in this study is based on a correct, although over-parameterized, model and is expected to behave like the χ^2 difference test if the true zero parameters are included in the test procedure.

The z Test: The z test provides an alternative procedure to the backward search procedure for dropping parameters. It can be implemented as a one-time rather than incremental procedure. Defining Θ_z as the set of free parameters that have nonsignificant z-test values, the traditional approach to dropping parameters simply involves eliminating the set Θ_z. It can be anticipated that the z-based procedure will drop more parameters than the W-based procedure, because the Wald test in this study was carried out only on the parameters in Θ_z. Since the χ^2 increment will be the stopping criterion, the Θ_W will always be a subset of Θ_z. Consequently, the W test provides a more conservative approach than the traditional z test. It remains to be seen whether the W test procedure will be too conservative, or the z test procedure too liberal.

2.2. *Initial models for backward search*

To investigate the performance of backward search strategy in SEM, we used the examples (Models A and B) in MacCallum's (1986). Model A has three independent factors (F1, F2, and F3) and two dependent factors (F4 and F5) with two indicators for each factor. Model B contains two independent factors and four dependent factors. Again each factor has two indicators. To start the backward search procedure, one initial model was used for each of Models A and B. It should be noted that each initial model is a more general model than its corresponding true model. Consequently, the initial models, although containing more parameters than necessary, can be considered correct models.

The initial models developed in this study are assumed to have saturation on the structural models. In most standard SEMs, a model has two main parts. One part is the *measurement model*, which consists of the relationships among measured and latent variables. The parameters of the measurement model consist of error variances and factor loadings of measured variables on latent factors. The other part is *structural model*, which deals specifically with relations among latent factors. In order to be comparable with MacCallum's examples, this study assumes that the measurement model is adequate and that misspecification may exist only in the structural relations among latent variables. Consequently, the measurement model must have no fundamental misspecifications such as would be picked up by a pseudo-χ^2 test (Bentler and Bonett, 1980). If this condition were not met, of course, model modification would also have to address the measurement model. Correct specification of the measurement model needs to be guided by substantive theory. Empirically, deciding the adequate number of factors is critical in finding an appropriate measurement model.

With an adequate measurement model, attention needs to be focused on the structural relations among latent variables. The backward search strategy starts with the saturated structural model in which all possible pairwise relations among latent factors are fully specified. In general, such a specification will include covariances among predictor factors and regression coefficients between predictor factors and outcome factors, although in a confirmatory factor analysis only factor covariances would be involved. The question to be answered is whether backward search can correctly locate the parameters needed to be retained and also drop the parameters that are not needed. Restricted search strategies were designed for each initial model during the backward search process as explained below.

The true Model A in MacCallum (1986) with nonzero parameters in the structural model is presented in Fig. 1a. The initial model, $M_0(A)$, for backward search based on MacCallum's (1986) true Model A was developed to contain a saturated structural model (see Fig. 1b). The solid lines or curves depicted in Fig. 1b represent those specified in the true Model A in MacCallum's (1986) . The dotted line from F3 to F5 is the only additional parameter that was included in the initial model $M_0(A)$. The following standard notation will be used: Φ is the covariance matrix for independent factors; Γ is the regression weight matrix of dependent factors on independent factors; and B is the corresponding matrix among dependent factors. Two different W-test procedures were implemented on $M_0(A)$; i.e., two sets of parameters were considered when executing the Wald test. The first procedure restricted the search for parameters to drop to be only in the Γ and B matrices. The second procedure also considered parameters in Φ in addition to those in Γ and B. We shall refer to these two procedures as dealing with parameter Set 1 (Γ, B) and Set 2 (Φ, Γ, B), respectively. Note that to obtain the true model only the path from F3 to F5, which is a parameter in the Γ matrix, can be dropped in the backward search procedure. Dropping any free parameters in Φ or B will result in errors, since these parameters are needed in the true model.

The initial model, $M_0(B)$, designed for the true Model B in MacCallum (1986) (see Fig. 2a) with both the Γ and B matrices saturated is presented in Fig. 2b. This include all the possible paths from the independent factors, F1 and F2, to the dependent factors, F3, F4, F5, and F6, and that among dependent factors (the dotted lines in

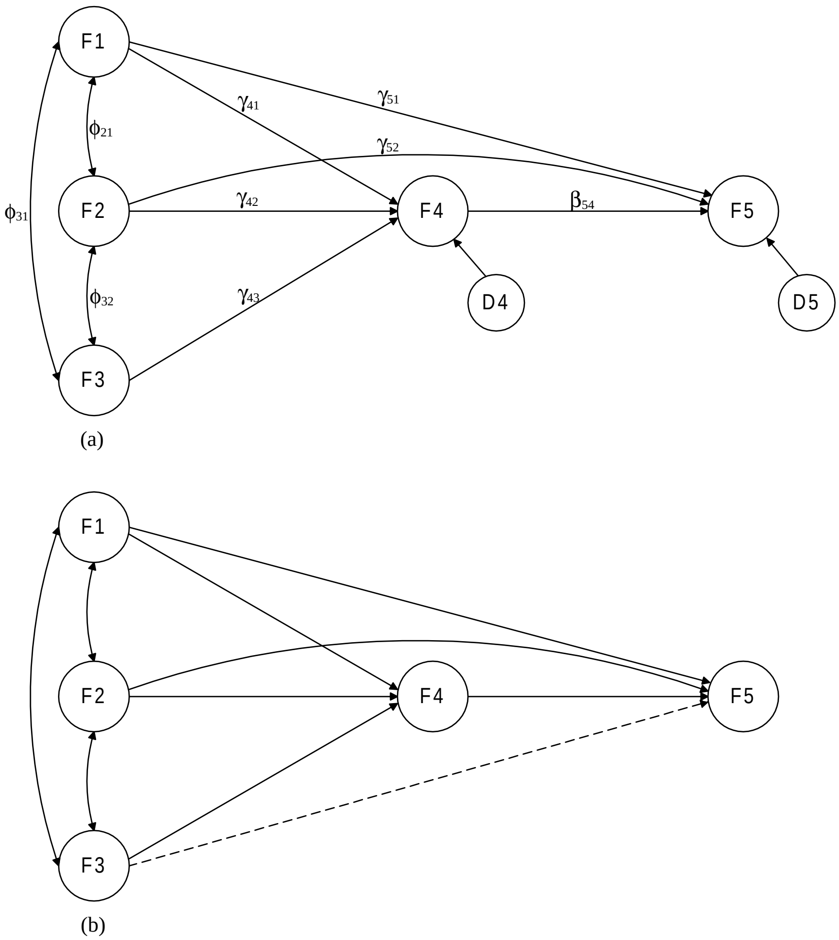

Fig. 1. (a)True Model A (MacCallum, 1986). (b) Model A with saturated structural model. Residual variables not shown.

Fig. 2b indicated the additional free parameters in Γ and B from the true Model B). The B matrix is a lower triangular matrix with diagonal zeros since recursiveness of the model was assumed as in MacCallum's forward search. This assumption, which was also made for Model A, is quite critical to this study since it indicates that the causal sequence of latent factors has been determined a priori. Three restricted search procedures for the initial model were implemented. The first search was restricted only to all the elements in the Γ matrix, the second only to those in the B matrix, while the last targeted all the elements in both Γ and B matrices.

For each true model, 100 samples with 300 cases in each sample were generated. Maximum likelihood estimates and the Wald-test procedure in EQS (Bentler, 1995) were used for the analyses. Comparisons were made among the alternative starting

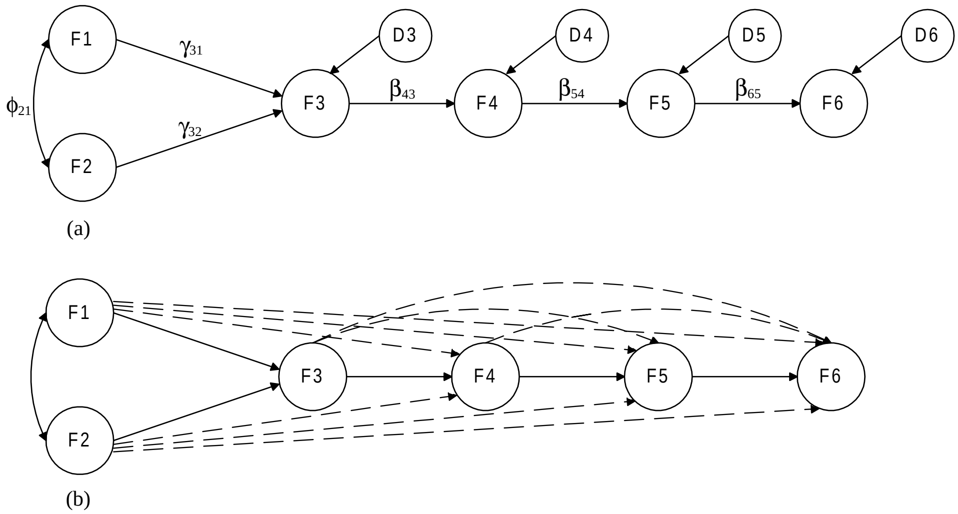

Fig. 2. (a) True Model B (MacCallum, 1986). (b) Model B with saturated Γ and B matrices. Residual variables not shown.

models and model modification procedures, as well as with the results of MacCallum's (1986) study.

3. Results

3.1. Population analyses

To adequately perform a backward search on the saturated structural model, two assumptions on model specification were made in this study. The first one is that the measurement model is correct, and the other is that the order of factors was correctly determined. While the latter should be based on substantive theory and can not be adequately evaluated with the Monte Carlo study, the former can be easily tested using a saturated confirmatory factor analytical (CFA) model. Using the hypothesized factors developed in each of the true model, we developed a saturated CFA model for each initial model and evaluated it with population analyses where the population covariance matrix was analyzed. The number of parameters associated with the saturated CFA model, therefore, is the same as that in the initial models. Each initial model and its corresponding saturated CFA model can be considered as equivalent models (Lee and Hershberger, 1990; Luijben, 1991; Stelzl, 1986) which yielded the same χ^2-test statistics.

We then conducted a Wald test on all the factor covariance parameters in the saturated CFA model to investigate if any of these parameters should be dropped. For the saturated CFA model based on $M_0(A)$ (see Fig. 1b), the population analysis yielded a perfect fit with 25 degrees of freedom (df) as expected. This provides an indication that the measurement model is perfectly fitted and is not affected by the specification of the structural model. Furthermore, the parameter estimates for the factor loadings and error variances associated with the measurement model are the same as the true values. For initial model $M_0(A)$, the Wald test correctly indicated that path $F3 \rightarrow F5$ should be excluded from the model. The Wald test using the saturated CFA model, however, indicated that all the parameters associated with factors are significantly different from 0 and should be kept in the model. Similar findings were obtained from comparing $M_0(B)$ and its corresponding saturated CFA model. Parameters associated with the measurement model were also correctly estimated, while parameters associated with the structural model showed a less conclusive pattern. Unlike the results obtained from $M_0(A)$, the results for $M_0(B)$ showed that several correlation coefficients among factors were not significantly different from zero.

3.2. Model A

The starting model for Model A, $M_0(A)$, was saturated in all Φ, Γ, and B matrices such that all parameters in these matrices were estimated. In backward searching for Model A, two test procedures on two sets of free parameters were considered when executing the W test. In the first procedure on Set 1, the W test was carried out only on the free parameters in Γ and B. More specifically, the parameters tested were γ_{41}, γ_{42}, γ_{43}, γ_{51}, γ_{52}, γ_{53}, and β_{54}. In Set 2, the Φ matrix was also involved. In other words, the W test evaluated the parameters ϕ_{21}, ϕ_{31}, and ϕ_{32} in addition to those in Γ and B. The Set 1 procedure can be considered as a more restricted specification search, since the parameters in Φ are determined to remain in the model whatever their statistical contribution may be. A correct backward specification search on either Set 1 or Set 2 in a single analysis for Model A would eliminate γ_{53} from the original model, and no other parameter. The primary measure from the study is the number of times out of 100 replications that the exactly correct original model was obtained with a particular searching strategy. Table 1 summarizes the results of backward search with $M_0(A)$.

The top part of Table 1 gives the results for searches in parameter Set 1 (Γ, B). The first two rows of this part summarize the frequencies for all the parameters that the z test and the W test suggested to drop. Each entry in these rows gives the number of replications out of 100 that a particular parameter was dropped by the given test. It should be noted again that results for the z test summarized frequencies for the parameters that are not significant, while results for the W tests summarized frequencies for the parameters that were not significant based on the χ^2 increment discussed earlier. It can be seen that the z and W tests correctly dropped γ_{53} 96 and 88 times, respectively. While the small number of misses (4 out of 100) can be expected by chance alone in the z test, the W test seemed to indicate that it is more conservative in dropping parameters. In addition, the z test also dropped parameters γ_{52} and β_{54} more frequently than expected. Parameter β_{54} fared the worst, being dropped 22 and 9 times by the z and W test procedures, respectively.

Part (b) on the top half of Table 1 provides the results in terms of correct determination of the complete models. The first row of part (b) shows that 64 out of 100 times

Table 1
Percentage of dropping a given parameter using z or W tests for initial model A (N = 300 with 100 replications)

	Parameter[a]										%[b]	
	ϕ_{21}	ϕ_{31}	ϕ_{32}	γ_{41}	γ_{42}	γ_{43}	γ_{51}	γ_{52}	$\underline{\gamma_{53}}$	β_{54}	z	W
Initial model of Fig. 1b (search restricted to Γ and B)												
(a) $z=0$	—	—	—	0	0	0	0	12	96	22		
$W=0$	—	—	—	0	0	0	0	8	88	9		
(b) Θ_z or $\Theta_W=0$												
1. $(\gamma_{53})=0$									d		64	75
2. $(\gamma_{53},\gamma_{52})=0$								d	d		10	5
3. $(\gamma_{53},\beta_{54})=0$									d	d	20	8
4. Other[c]											6	12
Initial model of Fig. 1b (search restricted to Φ,Γ and B)												
(a) $z=0$	2	5	1	0	0	0	0	12	96	22		
$W=0$	2	6	2	0	0	0	0	7	88	9		
(b) Θ_z or $\Theta_W=0$												
1. $(\gamma_{53})=0$									d		61	72
2. $(\gamma_{53},\gamma_{52})=0$								d	d		9	3
3. $(\gamma_{53},\beta_{54})=0$									d	d	17	8
4. Other[c]											13	17

[a]Parameter underlined, γ_{53}, has the true value of 0.
[b]% refers to percentage a set of parameters, Θ_z or Θ_W, is suggested to be dropped.
[c]Other indicates that parameter sets other than listed above are suggested to be dropped.
[d]Indicates the corresponding parameter being included in Θ_z or Θ_W.

the z test led to the true model in which only γ_{53} was correctly dropped, while the success rate for the W test was 75%. Row 2 indicates that 10 and 5 times, γ_{52} along with γ_{53} were eliminated by the z and W tests, respectively. Row 3 shows that β_{54} along with γ_{53} were dropped 20% and 8% of time by the z and W tests, respectively. Rows 2 and 3, combined, thus refer to samples for which one error was made. The error consists of dropping true nonzero parameters. Examples of row 4 are mainly the samples in which no parameters were dropped, i.e., the true zero γ_{53} was retained as nonzero.

In Set 2, the parameters in Φ were also considered in both z and Wald tests. The results are presented in the lower half of Table 1. Again, β_{54} was eliminated from the model more frequently than expected by chance, and the error of dropping γ_{52} occurred somewhat more frequently than expected, as was the case in Set 1. In addition, as would be expected, elements of Φ also were dropped, but only to the extent expected by chance. Because more parameters were involved in Set 2, this less restricted search had a lower success rate in finding the true model than the previous one. Part (b) at the bottom of Table 1 showed that the success rate of accepting $\gamma_{53}=0$ was 61% with the z test and 72% with the W test. An extra parameter in addition to γ_{53} was also suggested to be eliminated in a total of 17 samples with the z test and only 8 samples with the W test. The success rates of the backward search with the z or *Wald* tests are considerably higher than that obtained by two of the three initial models for Model A that MacCallum (1986) developed for the forward search. Using a restricted search with the third initial model, MacCallum reached a success rate of 55% with a more intensive approach in which the forward search continued even though a nonsignificant goodness-of-fit test statistic had been obtained.

The success rates for both tests on Model A were substantively improved when an additional assumption that amounts to using a still further restricted search in which the parameter β_{54} is required. The result of such a search is seen when row 3 in both sets is combined with row 1. The success rates are then increased from 64% to 84% with the z test and from 75% to 83% with the W test as shown in the top half of Table 1. The success rates also improve from 61% to 78%, and 72% to 80% in the bottom half, for the z and W tests, respectively. This more restricted search reflects a 30% and 10% improvement in finding the true model using the z and W tests, respectively. In general, the W test with success rates higher than 70% performs better than the z test with success rates at 60% on both sets of backward search summarized in Table 1.

3.3. Model B

Model B contains two independent latent factors and four dependent factors. The true model (Fig. 2a) has two free parameters in Γ (γ_{31} and γ_{32}) and three in B (β_{43}, β_{54}, and β_{65}). The initial model, $M_0(B)$, that is saturated in both Γ and B matrices for backward specification search, is presented in Fig. 2b. Three sets of parameters were considered in three separate search procedures. The first search concentrates only on the set of parameters in Γ, the second search on the B, while the last search focuses on both Γ and B. The objective of the backward search is to drop $\Gamma_W = (\gamma_{41}, \gamma_{42}, \gamma_{51}, \gamma_{52}, \gamma_{61}, \gamma_{62})$ in the first procedure, $B_W = (\beta_{53}, \beta_{63}, \beta_{64})$ in the second procedure, and both Γ_W and B_W in the last procedure. It should be noted that either of the first two search procedures, referred to as Γ- and B-restricted searches, respectively, contained extra parameters not needed in the model. The Γ-specific search did not consider the extra parameters specified in the B matrix, while the B-specific search does not consider those extra parameters in the Γ matrix. The Φ matrix containing only one parameter in $M_0(B)$ is not considered in the search for the true Model B.

The results of backward elimination with these three procedures are presented in Table 2. As before, the table was divided into segments. Each segment summarized results of each search procedure. Within each segment, part (a) reports the results for individual parameters and part (b) refers to the results on complete final models that were achieved. The results based on the z and W tests for each parameter (reported in part (a)) in the first search (the Γ-restricted searches) were very good indeed, with more than 95% of the time in correctly dropping true zero parameters and virtually no dropping of true nonzero parameters. The results on the complete model were similar to those for Model A. Success rates of 92% and 88% in dropping the right parameters by the z and W tests, respectively, were obtained. There are six possible types of

Table 2
Percentage of dropping a given parameter using z or Wald tests for Model B ($N=300$ with 100 replications)

	Parameter[a]														%[b]	
	γ_{31}	γ_{32}	γ_{41}	γ_{42}	γ_{51}	γ_{52}	γ_{61}	γ_{62}	β_{43}	β_{53}	β_{54}	β_{63}	β_{64}	β_{65}	z	W
Initial model $M_0(B)$ (search restricted to Γ)																
(a) z test	1	0	98	99	100	98	99	99	—	—	—	—	—	—		
W test	1	0	99	99	100	96	98	98	—	—	—	—	—	—		
(b) Θ_z or $\Theta_W=0$																
1. $(\gamma_{41},\ \gamma_{42},\ \gamma_{51},\ \gamma_{52}\ \gamma_{61},\ \gamma_{62})=0$			d	d	d	d	d	d							92	88
2. $(\gamma_{41},\ \gamma_{42},\ \gamma_{51},\ \gamma_{52},\ \gamma_{61},\quad)=0$			d	d	d	d	d								1	2
3. $(\gamma_{41},\ \gamma_{42},\ \gamma_{51},\ \gamma_{52},\quad,\gamma_{62})=0$			d	d	d	d	d								1	2
4. $(\gamma_{41},\ \gamma_{42},\ \gamma_{51},\quad,\gamma_{61},\ \gamma_{62})=0$			d	d	d	d	d								2	2
5. $(\gamma_{41},\ \gamma_{42},\quad,\gamma_{52},\ \gamma_{61},\ \gamma_{62})=0$			d	d	d	d	d								0	0
6. $(\gamma_{41},\quad,\gamma_{51},\ \gamma_{52},\ \gamma_{61},\ \gamma_{62})=0$			d	d	d	d	d								1	2
7. $(\quad,\gamma_{42},\ \gamma_{51},\ \gamma_{52},\ \gamma_{61},\ \gamma_{62})=0$			d	d	d	d	d								2	1
8. Other[c]			d	d	d	d	d								1	3
Initial model $M_0(B)$ (search restricted to B)																
(a) z test	—	—	—	—	—	—	—	—	6	100	12	99	99	12		
W test	—	—	—	—	—	—	—	—	7	96	6	96	97	7		
(b) Θ_z or $\Theta_w=0$																
1. $(\beta_{53},\ \beta_{63},\ \beta_{64})=0$										d		d	d		70	71
2. $(\beta_{53},\ \beta_{63},\ \beta_{64},\ \beta_{43})=0$									d	d		d	d		6	7
3. $(\beta_{53},\ \beta_{63},\ \beta_{64},\ \beta_{54})=0$										d	d	d	d		10	4
4. $(\beta_{53},\ \beta_{63},\ \beta_{64},\ \beta_{65})=0$										d		d	d	d	11	7
5. Other[c]															3	11
	Hypothesis															
	γ_{31}	γ_{32}	γ_{41}	γ_{42}	γ_{51}	γ_{52}	γ_{61}	γ_{62}	β_{43}	β_{53}	β_{54}	β_{63}	β_{64}	β_{65}	z	W
Initial model $M_0(B)$ (search restricted to Γ and B)																
(a) z test	1	0	98	99	100	98	99	99	6	100	12	99	99	12		
(b) Θ_z or $\Theta_W=0$																
1. $(\Gamma_W, B_W)=0$			d	d	d	d	d	d		d		d	d		63	65
2. $(\Gamma_W, B_W$, except $1\theta)=0$															7	15
3. $(\Gamma_W, B_W$, except $2\theta's)=0$															0	4
4. $(\Gamma_W, B_W$, except $3\theta's)=0$															0	1
5. $(\Gamma_W, B_W)=0$ plus $1\theta=0$															27	12
6. Other[c]															3	4

[a]Parameters underlined, γ_{41}, γ_{42}, γ_{51}, γ_{52}, γ_{61}, γ_{62}, β_{53}, β_{63}, and β_{64}, have the true value of 0.
[b]% refers the percentage a set of parameters, Θ_z or Θ_W, is suggested to be dropped.
[c]Other indicates that parameter sets other than listed above are suggested to be dropped.
[d]Indicates the corresponding parameter being included in Θ_z or Θ_W.

final models with a single error on not dropping a correct parameter in Γ. The results showed that from 1% to 2% of the runs yielded such a result as is shown in rows 2–7 under part (b).

The results of runs with a B-restricted search, shown in the second part of Table 2, were worse than that in the Γ-restricted search. As seen in part (a), individual parameters that should have been correctly dropped generally were close to the expected level of 95%, at 96–100%. However, the parameters that should have been retained were more frequently dropped than expected by the z test (6–12%), while the W test dropped only 6% or 7% of the time on those parameters. Overall, the parameters in the B matrix were correctly dropped at 70% and 71% of the time by the z and W tests, respectively.

An unrestricted or "blind" backward search was performed in the final procedure, as reported in the bottom part of Table 2. Individually the parameter tests perform about as expected, as shown in part (a). The z test yielded 98–100% correct decisions on dropping the parameters not needed while the W test yielded 95% to 98% correct decisions considering all the parameters in both Γ and B matrices. This performance can be an indication of the accuracy of identifying the precisely correct model. As seen in part (b) of the bottom panel, there was a success rate of 63% by the z test, and a slightly higher rate of 65% for the W test. Both the z and W tests performed about the same in searching for the correct model with $(\Gamma_W, B_W) = 0$. Row 2 indicates that there are 7 and 15 samples in which one parameter in (Γ_W, B_W) was not correctly eliminated by the z and Wald tests, respectively. Similarly, rows 3 and 4 showed the percentages in which 2 and 3 parameters in (Γ_W, B_W) were incorrectly evaluated to be different from zero. Row 5 shows that one true nonzero parameter was suggested to be dropped in 27% and 12% of the cases according to the z and W tests, respectively. Again, since the W test carried out in this study is more conservative than the z test, it can be expected that the W test will usually drop fewer parameters than the z test. This effect can be found by considering the fifth row of part (b). Compared to the success rates for unrestricted and restricted forward searches at 20% and 55%, respectively, in MacCallum's 1986 study, results for Model B in this study showed that the backward search is a more promising procedure. It is also evident from the results that the z test is in general even more reliable than the W test in finding the true model.

4. Discussion

The similarity of the two specification search procedures in SEM and the forward and backward stepwise regression analysis allows us to learn from the empirical comparisons of the two selection procedures used in the regression modeling technique. Regression analyses with a forward stepwise approach have demonstrated that an earlier introduced variable may be "locked in" the model but subsequently turned out to be nonsignificant as new models are developed (Kerlinger and Pedhazur, 1973). This forward selection procedure, therefore, is dependent on the variables already selected in the model. The backward selection procedure starts with a model containing all variables and drops the ones that are not significant. Consequently, the backward procedure does not "lock" nonsignificant covariates in the model. Comparing the forward and backward model selection in regression analyses, Montel (1970) also illustrated a situation in which a forward search may overlook an excellent model but not the backward search. The application of a backward search in SEM has recently been proposed by Green et al. (1999).

MacCallum's (1986) study on specification searches offered some instrumental advices for researchers utilizing the technique of structural equation modeling. Among them the concept of restricted search (see also Silvia and MacCallum, 1988) is one of the most constructive suggestions since it urges researchers to concentrate more on the substantive and theoretical information and focus on meaningful model modifications. Another important suggestion is that forward specification searches should not be terminated as soon as a nonsignificant goodness-of-fit χ^2 is obtained. Although this advice contradicts what most researchers do when conducting structural equation modeling, it is an important rule that has proven to be essential to a successful search (MacCallum, 1986). MacCallum also noted that a successful search using the forward approach depends highly on the initial model selected. Unfortunately, there is no criterion to determine if one initial model is better than the other. In fact, the forward search "may work least well when they are needed most" (MacCallum, 1986, p. 118). More recent application of Bayes factor (Kass and Raftery, 1995) may provide an adequate alternative for the forward search in SEM (Lee and Song, 2002; Song and Lee, 2001).

This study proposed another approach to model modification. Instead of adding free parameters, or removing constraints, as in a forward search, this approach drops free parameter, or imposes constraints, and may, therefore, be considered as a backward search. The backward search needs to start from a more general model while the forward search starts with a more limited model. This study focused on the saturation of a structural model. The advantage of using a saturated structural model as a starting point is that it may help identify the portion of the model where modification is needed, either the measurement model or structural model. Basically, with a saturated structural model, a nonsignificant goodness-of-fit χ^2 indicates that modification may concentrate on the structural part of the model only, while a significant χ^2 implies that the measurement model also needs to be reconstructed. Two approaches can be taken to modify the measurement model: one is to retain the current structural model and to add more factor loadings; the other is to change the model, such as, by adding latent factors. These issues were not addressed in this study. It should be noted, however, that specifying a saturated structural model may require strong theoretical knowledge such as whether a latent factor is exogenous or endogenous, and the causal order of latent factors.

To be comparable to the findings of MacCallum's (1986) forward search, this study used his examples to study the performance of a backward search. Assumptions made in MacCallum's models are also hypothesized in this study. Among them, the strongest one was that the measurement model was assumed to be correct. Like MacCallum (1986), we also focused solely on the modification of structural models. Using restricted search strategies, we also assumed that the causal sequence among factors is known which is consistent with MacCallum's assumption of the recursiveness of the model. Two backward searching procedures were considered. The first one is the z test which evaluates if a specific parameter is significantly different from zero. The z testing procedure is the same as the univariate Wald test. The second backward searching procedure used the multivariate Wald test on the parameters with nonsignificant z values. It was carried out by including one parameter in the testing procedure at a time, and evaluating the χ^2 increment to indicate the contribution of the parameter to the multivariate Wald statistic. Using a significant χ^2 increment as the criterion to stop the backward search, the multivariate Wald test procedure will always eliminate less parameters than the z test, or at most, the same parameters as the z test. Consequently, the multivariate Wald test in this study can be considered as a more conservative approach that yields more fully parameterized models.

Results of Model A indicated that both z and W tests performed very well for both types of restricted search. The first search, which only included parameters in Γ and B matrices, could be considered as a more restricted search than the second, in which parameters in Φ were also evaluated. As would be expected from MacCallum's study, the results showed that higher success rates were yielded by the more restricted search procedure. If knowledge were available, another feasible assumption that a specific parameter (β_{54}) should not be equal to zero could be made. In this case, further restricted searches lead to even higher success rates.

Restricted searches for Model B also had a fairly high chance of excluding the unnecessary parameters. The first two test procedures were designed to search either saturated Γ or B matrices. It should be noted that the initial model specified is correct although it contains more parameters than needed. The relatively lower success rates in searching the parameters in B seems to indicate that parameters that are not needed in the earlier segments of causal relationships of a model are more likely to be detected. The final procedure of Model $M_0(B)$ illustrated an unrestricted backward search which considered all parameters in the structural model. Both z and W tests yielded success rates higher than the results reported in MacCallum's forward search (1986) with 63% and 65%, respectively.

In this study there was a confounding between the simultaneous vs. stepwise nature of the testing procedure and the type of test. The z test used information from a single general starting model only, and all parameters that might be modified were determined simultaneously from this starting model. This is not the only way the z test could be implemented. It could also be implemented sequentially in a stepwise manner, for example, by picking the one least significant parameter to eliminate, then reestimating the model, thereafter again picking the least significant parameter to eliminate, and so on. Whether this alternative z-test procedure would work better or worse in practice remains to be evaluated by another study. The W test procedure, like the z test, was implemented so as to use information from a single estimation run on a general starting model. On the other hand, the process was stepwise since parameters were evaluated for dropping in a sequential way (Chou and Bentler, 1996). The stepwise approach can be extended to a test procedure by reestimating the model after dropping certain parameters. For example, a repeated application of the W-test procedure, such as dropping Γ parameters before reevaluating the new model to drop B parameters, can be conducted in practice. The process should work better for the third search procedure of Model B which considered both Γ and B parameters in the same model. It may be expected that the repeated reestimation with dropping of only one parameter at a time with the z test may also work better in practice. However, such model restorations are extremely time consuming, and were not studied further because of their impractical nature.

There is a theoretical reason to expect forward stepping and backward stepping to work differentially in large samples. Consider a sequence of nested models M_i, M_j, M_k, starting with the most restricted model and going to the most general model. Assume that M_j is the true model. Forward model modification starts with model M_i and tries to find M_j. But the Lagrange Multiplier statistic that is used to find this more general model is in fact not central χ^2 distributed under either the null or alternative hypothesis for parameters that are truly zero or those that are truly nonzero. Thus the procedure may have difficulty in locating the correct parameter to modify. On the other hand, the Wald statistic, based on model M_k, is central χ^2 distributed only for those parameters that are truly zero so that a large χ^2 should be a clear cut clue (within sampling error) regarding parameters to drop.

One concern on the backward search proposed in this study is that the initial models specified were overparameterized which may lead to the nonconvergence in the estimation procedure. This problem, however, has not been encountered in the 200 samples simulated in this study.

There are limitations of the backward search approach proposed in this study. The major drawback is that the assumptions that the measurement model is correct and the sequence of latent factors is known. Both assumptions require strong theoretical support. While the former can be evaluated using a saturated CFA model, the latter depends totally on solid understanding of substantive theory. Although the latter assumption cannot be empirically tested, it should not be considered as a weakness of the backward search. Rather, it encourages more careful thinking of the causal relationships among factors. As demonstrated in the population analyses, using either a saturated CFA or a saturated SEM as the initial model, although both yielded the same χ^2 goodness-of-fit test statistics, may lead to different backward search outcomes. Examination of equivalent models (MacCallum et al., 1993) clearly requires more attention in constructing the initial models. Constructing adequate initial models also involves selection of adequate number of factors. Application of Bayes factor has also been recommended recently for this purpose (Lee and Song, 2002).

The generalization of the results of this study cannot be clearly determined. Under the conditions that the measurement model is correctly specified and the sequence of underlying factors is correctly determined the backward search did outperform the forward search using MacCallum's (1986) examples. However, the performance of the backward search requires more investigation with other models and other conditions. For example, only two indicators were assumed for each factor since the models were borrowed from MacCallum (1986). Model under-identification can become a problem in the estimation procedure. The sample size, among others, is one of the major concerns in obtaining reliable estimates in SEM. Both of the z and W statistics are functions of sample size. In this study, we have decided to use sample size of 300 based on MacCallum's findings. It is expected that success rates in finding the correct model will be lower with smaller sample size that has been observed in MacCallum (1986). It should be noted that the W test will be problematic when the sample size is small (Gonzalez and Griffin, 2001). With huge sample size, more parameters than expected will be dropped by these tests if backward search is needed. It will be beneficial to take the combination of sample sizes and complexity of models into consideration in future studies. The model modification procedure investigated in this study requires that models considered be nested within each other; however, it is not adequate for comparisons among nonnested models. Recent development of Bayes factor (Berger, 1985) can be applied to structural equation models for comparisons among nonnested models (Lee and Song 2001; Song and Lee, 2001).

Acknowledgements

The authors are greatly indebted to the Editor and reviewers for valuable comments. Production assistance of Dongyun Yang is appreciated.

References

Bentler, P.M., 1995. EQS: Structural Equations Program Model. Multivariate Software Inc., Los Angeles.

Bentler, P.M., Bonett, D.G., 1980. Significance tests and goodness of fit in the analysis of covariance structures. Psychol. Bull. 88, 588–606.

Berger, J.O., 1985. Statistical Decision and Bayesian Analyses. Springer, New York.

Chou, C.-P., Bentler, P.M., 1990. Model modification in covariance structure modeling: a comparison among likelihood ratio, Lagrange multiplier, and Wald tests. Multivar. Behav. Res. 25, 115–136.

Chou, C.-P., Bentler, P.M., 1996. Application of AIC with the Wald and Lagrange multiplier statistics: model modification in covariance structure analysis. Multivar. Behav. Res. 31, 351–370.

Gleman, A., Meng, X.L., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statist. Sinica 6, 733–807.

Gonzalez, R., Griffin, D., 2001. Testing parameters in structural equation modeling: Every "one" matters. Psychological Methods 6, 258–269.

Green, S.B., Thompson, M.S., Poirier, J., 1999. Exploratory analyses to improve model fit: errors due to misspecification and strategy to reduce their occurrence. Struct. Equation Modeling: A Multidisciplinary J. 6, 113–126.

Jöreskog, K.G., 1993. Testing structural equation models. In: Bollen, K.A., Long, J.S. (Eds.), Testing Structural Equation Models. Sage, Newbury Park, CA, pp. 294–316.

Jöreskog, K.G., Sörbom, D., 1988. LISREL 7: A Guide to the Program and Application. SPSS, Chicago.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Statist. Assoc. 90, 773–795.

Kerlinger, F.N., Pedhazur, E.J., 1973. Multiple Regression in Behavioral Research. Holt, Rinehart and Winston, New York.

Leamer, E.E., 1978. Specification Searches: Ad hoc Inferences with Nonexperiment Data. Wiley, New York.

Lee, S.Y., 1984. On testing function constraints in structural equation models. Biometrika 72, 125–131.

Lee, S.Y., Hershberger, S., 1990. A simple rule for generating equivalent models in covariance structure modeling. Multivar. Behav. Res. 25, 313–334.

Lee, S.Y., Song, X.Y., 2001. Hypothesis testing and model comparison in two level structural equation model. Multivar. Behav. Res. 36, 639–655.

Lee, S.Y., Song, X.Y., 2002. Bayesian selection on the number of factors in a factor analysis model. Behaviormetrika 29, 23–40.

Luijben, T.C., 1991. Equivalent models in covariance structure analysis. Psychometrika 56, 653–665.

MacCallum, R.C., 1986. Specification searches in covariance structure analysis. Psychol. Bull. 100, 107–120.

MacCallum, R.C., Roznowski, M., Necowitz, L.B., 1992. Model modification in covariance structure analysis: the problem of capitalization on chance. Psychol. Bull. 111, 490–504.

MacCallum, R.C., Wegener, D.T., Uchino, B.N., Fabrigar, L.R., 1993. The problem of equivalent models in applications of covariance structure analysis. Psychol. Bull. 114, 185–199.

Montel, N., 1970. Why stepdown procedure in variable selection. Technometrics 12, 591–612.

Satorra, A., 1989. Alternative test criteria in covariance structure analysis: a unified approach. Psychometrika 54, 131–151.

Silvia, E.S.M., MacCallum, R.C., 1988. Some factors affecting the success of specification searches in covariance structural modeling. Multivar. Behav. Res. 23, 297–326.

Song, X.Y., Lee, S.Y., 2001. Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. British J. Math. Statist. Psychol. 54, 237–263.

Steiger, J.H., Shapiro, A., Browne, M.W., 1985. On the multivariate asymptotic distribution of sequential chi-square statistics. Psychometrika 50, 253–263.

Stelzl, I., 1986. Changing a causal hypothesis without changing the fit; some rules for generating equivalent path models. Multivar. Behav. Res. 21, 309–331.