Journal of Modern Applied Statistical Methods

Volume 9 | Issue 1 Article 16

5-1-2010

# Can Specification Searches Be Useful for Hypothesis Generation?

Samuel B. Green
*Arizona State University*, samgreen@asu.edu

Marilyn S. Thompson
*Arizona State University*, m.thompson@asu.edu

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

## Recommended Citation

Green, Samuel B. and Thompson, Marilyn S. (2010) "Can Specification Searches Be Useful for Hypothesis Generation?," *Journal of Modern Applied Statistical Methods*: Vol. 9 : Iss. 1 , Article 16.
DOI: 10.22237/jmasm/1272687300
Available at: http://digitalcommons.wayne.edu/jmasm/vol9/iss1/16

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Journal of Modern Applied Statistical Methods
May 2010, Vol. 9, No. 1, 160-171

Copyright © 2010 JMASM, Inc.

1538 – 9472/10/$95.00

# Can Specification Searches Be Useful for Hypothesis Generation?

Samuel B. Green  Marilyn S. Thompson
Arizona State University

Previous studies suggest that results from specification searches, as typically employed in structural equation modeling, should not be used to reach strong research conclusions due to their poor reliability. Analyses of computer generated data indicate that search results can be sufficiently reliable for exploratory purposes with properly designed and analyzed studies.

Key words: Structural equation modeling, specification searches, Lagrange multiplier test, modification indices.

## Introduction

In specification searches, researchers seek to modify their hypothesized model by freeing or imposing model constraints. Particular constraints on model parameters are freed to maximize improvement in model fit or imposed to yield a more parsimonious model while minimizing loss of fit. The emphasis in searches is generally on freeing model constraints because researchers seek a better fitting model after determining their hypothesized model failed to closely fit the data. Accordingly, we focus our study on specification searches that relax model constraints.

Researchers may revise a model as a result of a specification search and, upon achieving good fit, describe strong conclusions about an obtained model or portray changes to the model as hypothesis generation. A specification search may be regarded as leading to strong conclusions if the results are used to create a single model and presented as a validation of that model. In contrast, a specification search is conducted for exploratory purposes if the results are used to generate one or more models that are presented as potential alternatives to the originally postulated model(s), and these alternative models are described as requiring validation in future research.

Search methods used to reach strong conclusions should have to meet very stringent psychometric standards. By contrast, standards for hypothesis generation might be more relaxed, but should be sufficiently rigorous to prevent researchers from wasting time and energy investigating models based on non-replicable specification searches. Previous research suggests that search procedures are inadequate to reach strong conclusions (MacCallum, 1986; MacCallum, Roznowski, & Necowitz, 1992; Silvia & MacCallum, 1988). However, it is unclear whether searches can be useful for hypothesis generation in that standards have not been used that are consistent with this objective. The purpose of the study is to evaluate whether specification search methods can yield sufficiently accurate results to be used for the purpose of hypothesis generation.

This study considers only sequential specification searches, those that relax constraints one at a time, rather than nonsequential searches, such as Tabu, which attempt to determine combinations of parameters that would maximize model fit (e.g., Marcoulides, Drezner, & Schumacker, 1998). Sequential search methods are used almost exclusively in practice. In addition, this study evaluates searches that involve only adding

---

Samuel Green is a Professor in the School of Social and Family Dynamics. Email him at: samgreen@asu.edu. Marilyn Thompson is an Associate Professor in the School of Social and Family Dynamics. Email her at: M.Thompson@asu.edu.

parameters to models that were constrained in the original model to be equal zero. Thus, search methods in this paper are discussed in the context of adding parameters to models rather than relaxing linear constraints in general.

## Specification Search Methods

Researchers must define implicitly or explicitly a search family of parameters that potentially could be added to an initial model. Parameters should be included in a search family only if they are interpretable within the context of the research study. Specification searches that add parameters from the search family involve forward selection, backward selection, or a combination of forward and backward selection.

### Forward Selection

The most popular forward selection methods are based on the Lagrange multiplier (LM) test (Chou & Bentler, 1990; Jöreskog & Sörbom, 1989; Sörbom, 1989). The LM test statistic evaluates whether one or more constraints imposed on parameters should be maintained and is asymptotically distributed as a $\chi^2$, with degrees of freedom equal to the number of constraints being evaluated. The LM test is also called a modification index if it is used to evaluate a single constraint.

Two sequential approaches are generally available that use the LM test: the LM-respecified method and the LM-incremental method. In the first step of the LM-respecified method, the parameter in the search family with the largest univariate LM $\chi^2$ statistic is selected. At the second step, the model is respecified to include this parameter, and then, among the remaining parameters in the search family, the parameter is selected with the largest univariate LM $\chi^2$ statistic. At the third step, the model is respecified to include the parameter selected at the previous step, and the process continues. At any step, the search stops when the *p* value for the largest LM $\chi^2$ statistic is greater than the prescribed alpha (e.g., .05).

The LM-incremental method is similar to the LM-respecified method in that the parameter from the search family is added at each step that maximally increases model fit according to univariate LM tests. However, with this approach, the univariate LM $\chi^2$ statistics are incremental and are obtained by partitioning a multivariate LM $\chi^2$ statistic into single-df $\chi^2$ statistics. At step 1, the LM-incremental and the LM-respecified methods are identical. However, at step 2, the model is not respecified. Instead, multivariate LM $\chi^2$ statistics are computed for the addition of two parameters to the model: the parameter selected at step 1 plus each of the remaining parameters in the search family. An incremental univariate $\chi^2$ statistic can now be computed at step 2 for each of these remaining parameters: the multivariate LM $\chi^2$ statistic for the parameter selected at step 1 plus a remaining parameter in the search family minus the largest LM $\chi^2$ statistic from step 1. The parameter is selected with the largest univariate incremental LM $\chi^2$ statistic. The process continues until the *p* value for the largest LM univariate incremental $\chi^2$ statistic is greater than a prescribed alpha (e.g., .05).

### Backward Selection

Alternatively, stepwise backward approaches may be applied using the Wald test (Bentler, 1995; Chou & Bentler, 2002; Satorra, 1989), which is asymptotically distributed as a $\chi^2$. With these approaches, all parameters in the search family are added to a model at the beginning of the search process. Then, parameters in the search family are deleted one at a time such that loss of model fit is minimized.

Two backward selection methods are the Wald-respecified and the Wald-incremental methods. The distinctions between these approaches are similar to those between the LM-respecified method and the LM-incremental method. With the Wald-respecified method, the parameter with the smallest Wald $\chi^2$ statistic is selected at each step, and then prior to the next step, the model is respecified to exclude the selected parameter. With the Wald-incremental method, the model does not have to be respecified at each step in that univariate Wald tests are obtained by partitioning a multivariate $\chi^2$ statistic into single-df $\chi^2$ statistics.

### Combination Forward-Backward Selection

Other sequential procedures might involve both forward and backward searches (Green, Thompson, & Poirier, 1999). Analogous to stepwise regression analysis, parameters from the search family could be added and deleted at each step in the search process. However, such an approach might be considered too cumbersome by researchers unless automated by a software package. Alternatively, a two-stage search process could be employed. In the first stage, parameters in the search family are sequentially added based on the LM test; then, in the second stage, the added parameters from the first stage are sequentially deleted based on the Wald test. A two-stage search process could be used as an alternative to a backward search approach if the latter approach is not possible because the model is underidentified when all parameters in the search family are added to the model before backward deletion.

### Methods to Minimize Errors in Specification Searches

Traditionally $\chi^2$ statistics for individual LM and Wald tests have been evaluated at the .05 level in sequential searches, disregarding the number of conducted tests. Green, Thompson, and Babyak (1998) and Hancock (1999) have offered methods for controlling for Type I errors with multiple LM tests, while Green, Thompson, and Poirer (2001) have suggested a method with Wald tests.

The suggested methods are adaptations of either the Roy union-intersection method (Roy, 1953) or Holm's sequential Bonferroni method (Holm, 1979). However, controlling the probability of committing Type I errors across tests is problematic in that it increases the probability of committing Type II errors, failure to add parameters to the model that should be included. Consequently, methods for controlling Type I error rates are more likely to produce accurate results if large samples are employed to minimize Type II error rates.

Errors due to misspecification occur if a parameter improves model fit at a step in a search process, but would fail to improve fit if the model had been correctly specified. This type of error occurs even when a search is conducted on the population and, therefore, is distinguishable from an error due to sampling fluctuation. Based on past research, it is known that misspecification errors are less likely to occur if researchers have a relatively well-specified hypothesized model, have few irrelevant and many relevant parameters in their search family, and have large samples (MacCallum, 1986; MacCallum, Roznowski, Necowitz, 1992; Silvia & MacCallum, 1988). Accordingly, researchers must carefully construct not only the hypothesized model, but also the search family, based on best available theory and a thorough understanding of the empirical research literature. Errors due to misspecification also are less likely to occur with large samples in that the search process is more likely to avoid errors due to sampling fluctuation and thus yield better specified models. In addition, choice of search methods may have an effect on the likelihood of committing misspecification errors. Forward sequential search methods are by far most popular; however, backward sequential methods might yield better results (Green, Thompson, & Poirer, 2001; Chou & Bentler, 2002) in that the model is initially respecified to include all parameters in the search, reducing the likelihood of misspecification errors.

### Purpose of the Study

MacCallum, Roznowski, and Necowitz (1992) demonstrated convincingly that forward search methods are not sufficiently replicable to be useful for reaching strong conclusions. As part of a larger analysis, they conducted a number of searches on data collected from 3,694 research participants. They initially determined the first four parameters that were added to a model for the total sample. For the purpose of their analyses, the total sample could be considered the population. Next, for 8 different sample sizes, they determined the parameters added in the first 4 steps of a specification search for 10 replicate samples drawn from the total sample. They reported the added parameters for all searches in tabular form. For each sample size, the percent of times the same 4 modifications were made in the 10 replicate samples and the total sample was presented. As shown in Table 1, zero percent of the searches with *N*s of 100, 150, 200, and 250 yielded the same 4 modifications as were made in the total sample. Even for the largest sample size (*N* = 1,200), only 60% of the searches produced the same 4 modifications.

Based on these results and others, methodologists tend to view specification searches skeptically (e.g., Boosma, 2000; Breckler, 1990; MacCallum, et al., 1992). MacCallum (1995) offered the following advice about searches:

> …researchers must be concerned about use of the model generation strategy in practice. Users of this strategy must acknowledge that they are engaging in exploratory model development. There is not necessarily anything wrong with exploratory model development as long as it is acknowledged in practice that that is what is being done and that the outcome is a model that cannot be supported without being evaluated using new data. Serious problems arise when the model generation strategy is applied without any effort to attach substantive meaning to model modifications and when the resulting model is treated as if it has been confirmed because it fits the observed sample data well. The model generation strategy is a legitimate approach to model development if it is used responsibly, but such use seems to be the exception rather than the rule in much of the applied literature. (p. 34)

Searches should not have to meet as stringent criteria to be used for hypothesis generation, but nevertheless should demonstrate adequate psychometric quality. Based on the results presented by MacCallum, et al. (1992), we computed match statistics using a less stringent criterion. Matching was assessed for each selected parameter in a replicate sample rather than for the set of all four selected parameters. In other words, a match occurred if a parameter selected in a sample was the same as one of the four parameters selected in the total sample. For any sample size, the maximum number of possible hits using this definition of a match is 40 (= 4 parameters x 10 replicates) rather than 10.

As shown in Table 1, the hit percentages were 30, 38, 40, and 55 for sample sizes of 100, 150, 200, and 250, respectively. The percent of hits continued to increase from 60 to 82 as sample size increased from 325 to 1,200. These results suggest that specification searches may be insufficiently accurate even for exploratory analyses at smaller sample sizes (perhaps less than a sample size of 250 based on these results), but acceptable for this purpose at larger sample sizes.

The findings of previous studies indicate that specification searches should not be used for

Table 1: Summary of Results from MacCallum, et al. (1992) Study: Match Statistics between Parameters Selected in the 10 Replicate Samples and the First Four Selected Parameters in the Total Sample

| Sample Size | Percent of 4:4 All Matches[a] | Mean Percent of 4:4 Any Matches[b] |
|---|---|---|
| 100 | 0 | 30 |
| 150 | 0 | 38 |
| 200 | 0 | 40 |
| 250 | 0 | 55 |
| 325 | 10 | 60 |
| 400 | 20 | 65 |
| 800 | 40 | 68 |
| 1,200 | 60 | 82 |

[a] All 4 parameters selected in sample match all 4 parameters selected in population. These percentages were presented by MacCallum, et al. (1992); [b] Any of 4 parameters selected in sample match any of 4 parameters selected in population. We calculated these mean percentages based on the results of the searches reported by MacCallum, et al. (1992).

reaching strong conclusions (MacCallum, 1986; MacCallum, Roznowski, Necowitz, 1992; Silvia & MacCallum, 1988). However, it is less clear whether searches can be used to meet exploratory goals. This study analyzed the covariance matrix examined by MacCallum, et al. (1992) in their classic study of specification searches and, by drawing samples based on this covariance matrix, investigated whether search results can be sufficiently accurate to warrant their use for hypothesis generation.

The strength of the approach in this study is that the examined covariance matrix was based on real data, and this matrix was investigated in a well known study that led to negative conclusions about specification searches. Given these negative results, it is important to establish that specification search methods can be useful for exploratory purposes. If this can be established, then researchers might be encouraged to conduct further research on these popular search methods, even though this research requires methods for tracking all possible additions to models and thus is time-consuming.

A number of authors (Bentler & Chou, 1993; Chou & Bentler, 1993; Kaplan, 1990; Sörbom, 1989) have convincingly argued that researchers should evaluate not only the $\chi^2$ values associated with tests in searches, but also statistics that assess expected change in a parameter when that parameter is freed. In this paper we focus on the $\chi^2$ values and do not attempt to address the broader and more complex issue of combining the results of significance tests with the expected change parameter statistic.

## Methodology

### Initial Model and Data

Searches were conducted using the hypothesized model and covariance matrix examined by MacCallum, et al. (1992) in their study of specification searches. The heuristic model (presented in Figure 1) is of employee responses to affect (Hulin, Roznowski, & Hachiya, 1985) and includes 21 measured variables and 7 factors. The data are based on a questionnaire administered to 3,694 employees from two hospitals (see MacCallum, et al. (1992) for a more detailed description of the model and the data.) The raw data were not available. Consequently, the covariance matrix was treated as the population covariance matrix, and samples were generated based on this matrix using the multivariate normal generator available in EQS (Bentler, 1995).

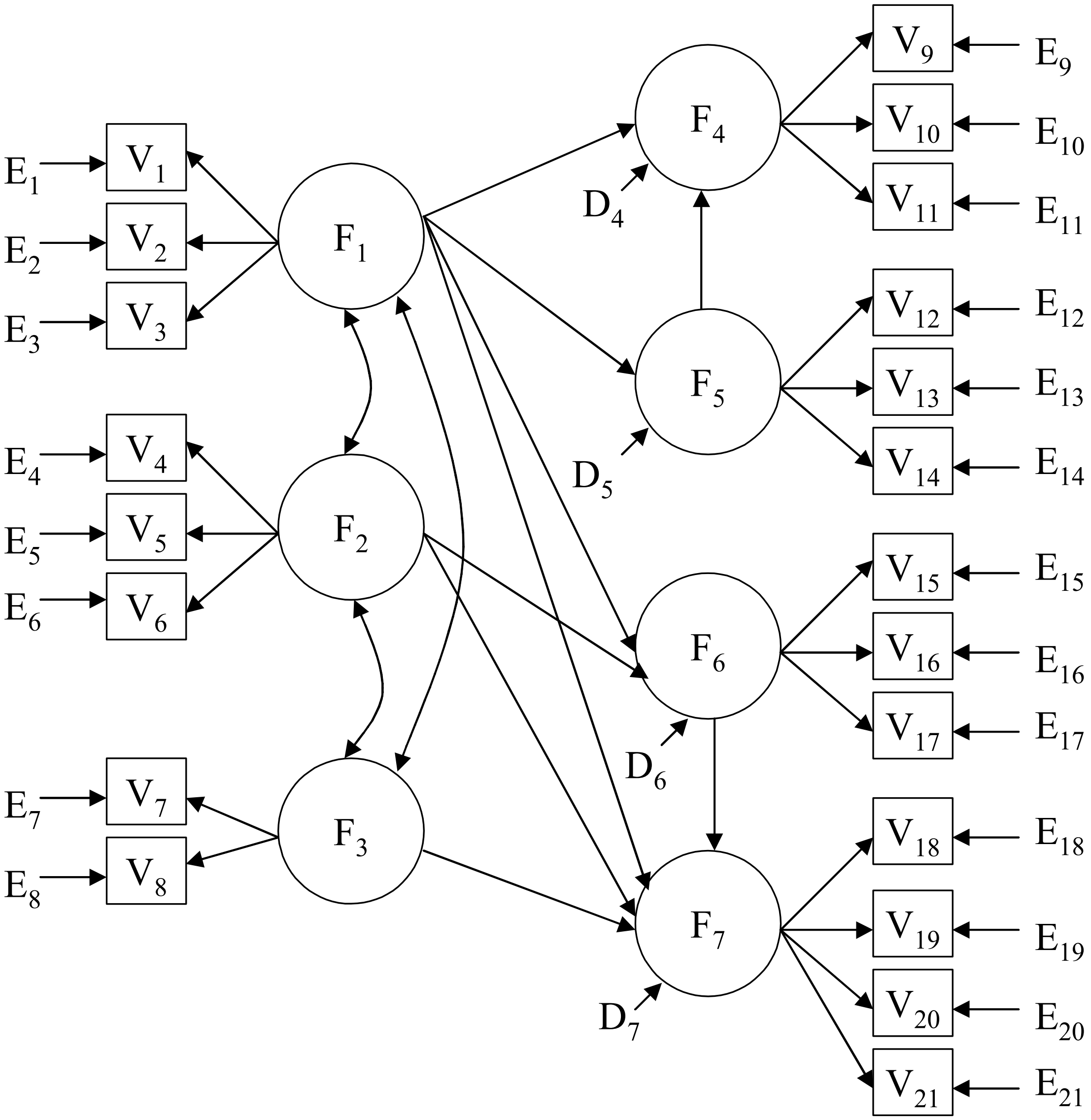
One thousand replicate samples were generated for each of four sample sizes: 200, 500, 800 and 1,200. If difficulties emerged in the estimation process (e.g., iterative process failed to converge; parameter estimates were out of bounds), additional replicate samples were generated to yield 1,000 replicates. The sample sizes were similar to those explored by MacCallum et al. (1992), which varied in size from 100 to 1,200. Sample sizes less than 200 were excluded because past research indicates that these sample sizes are inadequate for many applications of SEM (Tomarken & Waller, 2005) and, in particular, have been shown to be insufficient for specification searches (MacCallum, 1986; MacCallum et al., 1992; Silvia & MacCallum, 1988).

### Forward Searches

The LM-incremental method was used to conduct the searches. This method was chosen for four reasons: (a) the LM-incremental method is available within EQS (Bentler, 1995) and, accordingly, is presumably a relatively popular approach; (b) forward search procedures, such as the LM-incremental method, are most frequently applied and, in this sense, are more interesting to explore; (c) the LM-incremental method is efficient in that it does not require model respecification; and (d) the belief that the replicability of search results using the LM-incremental test should be similar to those using the LM-respecified method, the alternative forward approach for conducting specification searches.

For each sample size, two sizes of search families were considered to assess whether a decrease in family size would increase the replicability of specification search results. The small family contained 69 parameters. These parameters included (a) 16 paths between the exogenous factors $F_1$ through $F_3$ and their indicators $V_1$ through $V_8$, (b) 39 paths between the endogenous factors $F_4$ through $F_7$ and their indicators $V_9$ through $V_{21}$, (c) 5 paths between

Figure 1: Initially Hypothesized Model from MacCallum, et al. (1992)



Factors 1, 2, and 3 are work satisfaction, pay satisfaction, and perceptions about physical working conditions. Factors 4, 5, 6, and 7 are change to improve work conditions, citizenship such that individuals volunteer and display extra-role behaviors in the work place, psychological or passive withdrawal of individuals from the workplace, and thoughts and intentions about physical withdrawal from the organization in the future.

the exogenous factor $F_1$ through $F_3$ and the endogenous factors $F_4$ through $F_7$, and (d) 9 paths among the endogenous factors $F_4$ through $F_7$, excluding the path from $F_4$ to $F_5$ to avoid an underidentified model if selected. Covariances among residuals were ignored in the small family because they are likely in many applications to be included in models to improve model fit without conceptual justification (Hoyle & Panter, 1995). The large family contained 178 parameters and included (a) the parameters in the small family, (b) 27 covariances among the indicator errors $E_1$ through $E_8$, (c) 78 covariances among the indicator errors $E_9$

through $E_{21}$, and (d) 4 covariances among disturbances $D_4$ through $D_7$ associated with the endogenous factors. The covariance between $E_7$ and $E_8$, the covariance between $D_4$ and $D_5$, and the covariance between $D_6$ and $D_7$ were not part of the family in that freeing these parameters would have produced an underidentified model. The large family is similar in size to the one used by MacCallum, et al. (1992) in their searches.

MacCallum, et al. (1992) reported results for the replicability of the first four parameters added to the initial model based on searches. For their searches, all added parameters were significant at the .01 level or lower. This study examined both the first four parameters as well as the first eight parameters added to the model in sample data. At any step within a search, a parameter was not added to the model and the search was terminated if the parameter was not significant at the .05 level. The first four added parameters always were significant at the .05 and, most often, at a much lower value. For the next four added parameters, the search was terminated for some replicate samples with a sample size of 200 and the small family size. For this combination of conditions, the search was discontinued 37 times out of 1,000 replicate samples: 1 time at step 6, 7 times at step 7, and 29 times at step 8.

### Assessment of Replicability

Replicability was assessed by computing the extent to which the results of the LM-incremental method in the sample matched those in the population. An average match rate across the 1,000 replications was computed, with four different definitions for a match.

In order to reach strong conclusions based on specification searches, it would be ideal if all parameters selected to be added to the model in the sample to match those found in the population. To assess the accuracy of searches for this purpose, a stringent criterion was defined: the 4:4 All Match. For this criterion, all four parameters selected first in the sample had to be the same as all four parameters selected first in the population. For any one replication, the selected parameters either matched or failed to match. The percent of 4:4 all matches was computed across the 1,000 replications. This definition of percent of matches is the same as the one presented by MacCallum, et al. (1992).

If search methods are not used to reach strong conclusions, but rather to generate hypotheses, less stringent criteria can be used for matches. Three less stringent criteria were defined as follows:

- 4:4 Any Match: Any 1 parameter of the first 4 parameters selected in the sample matches any 1 parameter of the first 4 parameters selected in the population. The maximum number of matches for a replication was 4 and occurred if all 4 of the sample parameters matched all 4 of parameters selected in the population.
- 8:4 Any Match: Any 1 parameter of the first 8 parameters selected in the sample matches any 1 parameter of the first 4 parameters selected in the population. The maximum number of matches for a replication was 4 and occurred if any 4 of the 8 parameters selected in the sample matched the first 4 parameters selected in the population.
- 4:12 Any Match: Any 1 parameter of the first 4 parameters selected in the sample matches any 1 parameter of the first 12 parameters selected in the population. The maximum number of matches for a replication was 4 and occurred if the 4 parameters selected in the sample matches any 4 of the 12 parameters selected in the population.

For each of these three criteria, the percent of matches was computed for each replication (0%, 25%, 50%, 75%, or 100%), and the mean percent was computed across the 1,000 replications.

From a hypothesis generating perspective, the 4:4 any match (as well as the 4:4 all match) might be considered too stringent in comparison with the 8:4 any match. Researchers can choose to conduct more steps in the search process than the number of parameters that they actually add to their model. For example, they might continue the search process through the first 8 steps rather than the first 4 steps. By conducting this deeper search, they are more likely to find the first four parameters added in the population. To the extent that the mean percentage for 8:4 any matches exceeds the mean percentage for 4:4 any matches, deeper searches are recommended. The 8:4 any match is consistent with searches for hypothesis generation in that researchers are not likely to include all new parameters indicated by a search. Even if parameters are included in a search family only if they are conceptually meaningful, some combinations of parameters are more meaningful than others and thus make for better generated hypotheses.

Three of the four matching criteria assessed matches to the first four parameters added to the model at the population level. However, the choice of the number of added parameters was arbitrary. Even with 12 parameters added to the model, the fit was not perfect in the population. Accordingly, a criterion was included in the analyses matching sample searches to the first 12 parameters added to the model at the population level.

### Search Results at the Population Level

An appropriate initial model for this study should not fit so well that a search is unnecessary, but not so badly that a search would be fruitless because the model is severely misspecified. For the original sample data explored by MacCallum, et al. (1992; *N* = 3,694), the model (i.e., Figure 1) fit the data adequately, but not as close as desired by many researchers, $\chi^2$ (177) = 3215.44, *p* < .01, *CFI* = .89, *RMSEA* = .068 (90% CI of .066 to .070). The CFI is less than either the traditional cutoff of .90 or the cutoff of .95 recommended by Hu and Bentler (1999), while the RMSEA is slightly greater than the cutoff of .06 (Hu & Bentler, 1999). Consequently, a search would be considered by many researchers.

The model fit much better after the first four parameters were added based on LM-incremental method for either the small or the large search family (same four added parameters for both families), $\chi^2$ (173) = 1825.06, *p* < .01, *CFI* = .94, *RMSEA* = .051 (90% CI of .049 to .053); the model shows further improvement in fit if more parameters are added. Model fit improved when 4 more parameters were added to the model based on the LM-incremental method for the small search family (total of 8 added parameters): $\chi^2$ (169) = 1396.68, *p* < .01, *CFI* = .96*, RMSEA* = .044 (90% CI of 042 to .047). It increased even further with 4 more added parameters (total of 12 added parameters): $\chi^2$ (165) = 1177.21, *p* < .01, *CFI* = .96*, RMSEA* = .041 (90% CI of .039 to .043).

To the extent that the use of the LM-incremental and the LM-respecified methods yielded very different sets of added parameters, there is less confidence that the results of this study based on the LM-incremental method would generalize to those using the LM-respecified method. Of course, even if the results were identical, one could not be confident that the findings of this study would generalize to those based on the LM-respecification method. As shown in Table 2, the two methods produced similar – but not identical – results at the population level for both the small and large search families.

## Results

### Percentages Based on 4:4 All Matches

As shown in Table 3, the percentages for 4:4 all matches were uniformly low for the large search family, with the percentages varying from 6% for a sample size of 200 to 39% for a sample size of 1,200. The percentages for 4:4 all matches were higher for the small search family, but did not exceed 50% for sample sizes of 500 or less. The percentages approached 80% for only the largest sample size of 1,200.

### Percentages Based on Alternative Definitions of Matches

The percentages improved considerably when a match did not require all of the first four parameters in the sample to match all of the first four parameters in the population. For 4:4 any matches, the means for the percent of matches for the small search family ranged from 65% for an *N* of 200 to 94% for an *N* of 1,200. The means for the percent of matches were 12 to 13 percentage points lower for the large search family, ranging from 52% for an *N* of 200 to 81% for an *N* of 1,200.

For more liberal definitions of a match, the match percentages were higher under comparable conditions. The means for the percent of 8:4 any matches always exceeded 90% except with *N*s of 200. The means for the

Table 2: Ranks of Entry into Model Based on Incremental LM Test and LM Test with Model Respecification in the Population

| Search Family Size of 69 Parameters | | | Search Family Size of 178 Parameters | | |
|---|---|---|---|---|---|
| Parameters | Incremental-LM Method | Respecified-LM Method | Parameters | Incremental-LM Method | Respecified-LM Method |
| V17, F5 | 1 | 1 | V17, F5 | 1 | 1 |
| V19, F6 | 2 | 3 | V19, F6 | 2 | 3 |
| V20, F6 | 3 | 4 | V20, F6 | 3 | 4 |
| V11, F7 | 4 | 2 | V11, F7 | 4 | 2 |
| V3, F3 | 5 | 5 | E20, E19 | 5 | 13 |
| F5, F7 | 6 | 6 | V3, F3 | 6 | 5 |
| V12, F4 | 7 | 9 | E17, E12 | 7 | 6 |
| V20, F4 | 8 | --- | F5, F7 | 8 | 7 |
| V15, F5 | 9 | 7 | V12, F4 | 9 | 9 |
| V16, F7 | 10 | 11 | E19, E14 | 10 | --- |

Note: For path parameters, the second variable affects the first variable. Parameters are shown if they were selected in the first 10 steps using the LM-incremental method. Ranks are not presented if the rank of entry into the model exceeded 15 steps.

Table 3: Match Statistics between Parameters Selected in Sample and Population for Different Sizes of Samples and Search Families

| Mean Match Statistics | Sample Size 200 | 500 | 800 | 1,200 |
|---|---|---|---|---|
| Smaller Search Family Size: 69 Parameters | | | | |
| Percent of 4:4 all matches (All 4 parameters selected in sample match all 4 parameters selected in population.) | 17 | 46 | 67 | 78 |
| Mean percent of 4:4 any matches (Any of 4 parameters selected in sample match any of 4 parameters selected in population.) | 65 | 82 | 90 | 94 |
| Mean percent of 8:4 any matches (Any of 8 parameters selected in sample match any of 4 parameters selected in population.) | 82 | 94 | 98 | 99 |
| Mean percent of 4:12 any matches (Any of 4 parameters selected in sample match any of 12 parameters selected in population.) | 80 | 93 | 98 | 99 |
| Larger Search Family Size: 178 Parameters | | | | |
| Percent of 4:4 all matches (All 4 parameters selected in sample match all 4 parameters selected in population.) | 6 | 21 | 32 | 39 |
| Mean percent of 4:4 any matches (Any of 4 parameters selected in sample match any of 4 parameters selected in population.) | 52 | 70 | 77 | 81 |
| Mean percent of 8:4 any matches (Any of 8 parameters selected in sample match any of 4 parameters selected in population.) | 69 | 90 | 95 | 98 |
| Mean percent of 4:12 any matches (Any of 4 parameters selected in sample match any of 12 parameters selected in population.) | 63 | 79 | 84 | 86 |

percent of 4:12 any matches were relatively high when the family size was small; that is, they exceeded 90% as long as *N* was greater than 200. However, the means for the percent of 4:12 any matches were generally lower for the large family size, ranging from 63% to 86%.

## Conclusion

The results based on matching all of the first four parameters added in the sample with all of the first four parameters added in the population (i.e., 4:4 all matches) are consistent with those of MacCallum, et al. (1994) and do not support the use of search methods for reaching strong conclusions unless the size of the search family is quite restricted and sample size is very large. For the model in this study, the percent of matches approached 80% only if the sample size was 1,200 and the family size was small. Conversely, the results based on more liberal match criteria support the use of search methods for exploratory purposes with moderately large sample sizes and a small search family. The mean of the percentages for 8:4 any matches and 4:12 any matches were generally satisfactory (i.e., approximately 80% or greater) if sample size was 500 or greater. The mean percentages were always higher for the smaller search family.

Results demonstrate that specification searches can produce replicable results for exploratory purposes. However, the reported percentages are specific to the model and the data set examined. Although the sample size of 500 was adequate in a number of conditions for the searches conducted in our study, smaller sample sizes might be satisfactory if the initial model fails to include a minimal number of relevant parameters and the search is conducted on a very restricted search family. By contrast, larger sample sizes might be required if the initial model excludes a large number of relevant parameters and the search family is very large. It is suggested that other studies using real data sets and generated data be conducted to assess the generalizability of our results.

The findings in this study are consistent with previous recommendations about conducting specification searches (Green et al., 1999; Kaplan, 1990; MacCallum, 1986; MacCallum, Roznowski, & Necowitz, 1992; Schumacker & Lomax, 1996; Silvia & MacCallum, 1988). Researchers should (a) narrow the number of parameters in searches based on their conceptual understanding of the substantive area and the methods employed; (b) conduct searches on large samples; and (c) carry out deep searches in that parameters generated later in a search might be appropriate to add to an initial model. Researchers may be hesitant to carry out searches with small search families in that modified models resulting from searches are likely to yield better fit if more parameters are included in search families. However, this improvement in fit is likely to be illusory, resulting from overfitting a model to the peculiar characteristics of a specific sample; that is, the improved fit is unlikely to hold up in replicate samples.

Based on our results, specification searchers can be appropriate for exploratory purposes if used judiciously and can be reported as potentially a valid method for hypothesis generation. In agreement with others who have written about specification searches (e.g., Boomsma, 2000; MacCallum, 1995), researchers should describe explicitly in their research publication their initial model, the search family, the search method, and the conceptual meaning of the added parameters so that readers can evaluate appropriately the meaningfulness of their results. In addition, they need to indicate the importance of validating in new samples the models that are generated through the searches. Because some researchers are hesitant to report the search process involved in conducting their SEM analyses, reviewers of manuscript in which SEM is applied should ask authors to describe their initial model and to delineate the search family and the search method employed to obtain their final model(s).

We suspect the results in this study would have been similar if we had investigated the LM-respecified method and linear constraints other than model parameters being equal to zero However, future research should investigate this hypothesis. It would also be useful to assess other search methods (e.g., backward selection) as well as the effect of controlling for Type I error rate across multiple tests.

## Acknowledgement

The authors would like to extend our appreciation to Dr. R. C. MacCallum for sharing the covariance matrix used in his 1992 study with us.

## References

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Chou, C.-P. (1993). Some new covariance structure model improvement statistics. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models*. Newbury Park, CA: Sage.

Boomsma, A. (2000). Reporting on structural equation analyses. *Structural Equation Modeling*, *7*, 461-483.

Chou, C.-P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange Multiplier, and Wald tests. *Multivariate Behavioral Research*, *25*, 115-136.

Chou, C.-P., & Bentler, P. M. (2002). Model modification in structural equation modeling by imposing constraints. *Computational Statistics & Data Analysis*, *41*, 271-287.

Chou, C.-P., & Bentler, P. M. (1993). Invariant standardized estimated parameter change for model modification in covariance structure analysis. *Multivariate Behavioral Research*, *28*, 97-116.

Green, S. B., Thompson, M. S., & Babyak, M. A. (1998). A Monte Carlo investigation of methods for controlling Type I errors with specification searches in structural equation modeling. *Multivariate Behavioral Research*, *33*, 365-384.

Green, S. B., Thompson, M. S., & Poirier, J. (1999). Exploratory analysis to improve model fit: Errors due to misspecification and a strategy to reduce their occurrence. *Structural Equation Modeling*, *6*, 113-126.

Green, S. B., Thompson, M. S., & Poirer, J. (2001). An adjusted Bonferroni method for elimination of parameters in specification addition searches. *Structural Equation Modeling*, *8*, 18-39.

Hancock, G. R. (1999). A sequential Scheffé-type respecification procedure for controlling Type I error in exploratory structural equation model modification. *Structural Equation Modeling*, *6*, 158-168.

Hoyle, R. H., & Panter, A T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158-176). Thousand Oaks, CA: Sage Publications.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (*2nd Ed.*). Chicago: SPSS.

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, *25*, 137-115.

Marcoulides, G. A., Drezner, A., & Schumacker, R. E. (1998). Model specification searches in structural equation modeling using Tabu Search. *Structural Equation Modeling*, *5*, 365-376.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107-120.

MacCallum, R. C. (1995). Model specification, procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 16-36). Thousand Oaks, CA: Sage Publications.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490-504.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statistics*, *24*, 220-238.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, *54*, 131-151.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates

Silvia, E.S.M., & MacCallum, R.C. (1988). Some factors affecting the success of specification searches in covariance structural modeling. *Multivariate Behavioral Research*, *23*, 297–326.

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

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*, *1*, 31-65.