Recovery of weak factor loadings in confirmatory factor analysis under conditions of model misspecification

CARMEN XIMÉNEZ

Autónoma University of Madrid, Madrid, Spain

This article presents the results of two Monte Carlo simulation studies of the recovery of weak factor loadings, in the context of confirmatory factor analysis, for models that do not exactly hold in the population. This issue has not been examined in previous research. Model error was introduced using a procedure that allows for specifying a covariance structure with a specified discrepancy in the population. The effects of sample size, estimation method (maximum likelihood vs. unweighted least squares), and factor correlation were also considered. The first simulation study examined recovery for models correctly specified with the known number of factors, and the second investigated recovery for models incorrectly specified by underfactoring. The results showed that recovery was not affected by model discrepancy for the correctly specified models but was affected for the incorrectly specified models. Recovery improved in both studies when factors were correlated, and unweighted least squares performed better than maximum likelihood in recovering the weak factor loadings.

Factor analysis is one of the most widely used statistical procedures in psychological research. When conducting a factor analysis, the researcher must make a number of decisions that will have important consequences for the results obtained. For instance, the researcher must decide how many factors should be included in the model. In practical applications, researchers often face the problem of finding factorial structures containing one or more weak factors. A *weak factor* is one that shows relatively little influence on the set of measured variables or is defined by small loading sizes. A possible reason such factors could be present is the low reliability of the observed variables, which could be a consequence of inadequate wording of the items resulting in a high measurement error and a small percentage of common variance. In such cases, the variables at issue should be avoided. However, in other situations, estimating weak factors is important and the unreliability problem is unavoidable, because the items are well written. For instance, this could happen when measuring cognitive abilities or personality attributes that occupy a low position in the hierarchy of mental traits. One of the best known of such theories is Vernon's (1961) hierarchical group factor theory of the structure of human intellectual abilities, with Spearman's general factor (*g*) located at the top of the hierarchy and several major, minor, and specific group factors below *g*. This theory implies that most of the variance will be attributable to *g* and to the major factors, and that the contributions of the minor factors will be smaller. Among the major group factors is the verbal–numerical–educational factor, which splits into several

factors that vary from strong to weak (for more details, see Table V in Vernon, 1961, p. 23). In such cases, applied researchers must be aware of the consequences of working with factorial structures containing both strong and weak factors. Is the recovery of the weak factors adequate? Do all estimation methods recover the weak factors equally? Which conditions affect this recovery?

Previous research has especially addressed these issues in the context of exploratory factor analysis (EFA). For instance, in a simulation study that introduced model and sampling error, Briggs and MacCallum (2003) examined the performance of the maximum likelihood (ML) and unweighted least squares (ULS) estimation methods to recover a known factor structure with relatively weak factors. They found that in situations with a moderate amount of error and small sample sizes (e.g., $N = 100$), ML often failed to recover the weak factor, whereas ULS succeeded. In another study, MacCallum, Widaman, Preacher, and Hong (2001) examined the role of model error in the recovery of population factors in the context of EFA under varying conditions of sample size, number of factors, number of indicators per factor, and level of communalities for ML solutions. They found that, with high communalities and strongly determined factors, sample size had relatively little impact on the solutions, and good recovery of population factors could be achieved even with fairly small samples. However, sample size had a much greater impact as communalities entered the wide or the low range. More importantly, MacCallum et al. (2001) also found that, regardless of sample size, as long as the

C. Ximénez, carmen.ximenez@uam.es

model was correctly specified, model error did not influence the recovery of population factors.

Within the context of confirmatory factor analysis (CFA), the study of the recovery of weak factors possibly makes more sense because, in such models, the number of factors is specified in advance and the theoretical model may include both strong and weak factors. However, a more limited number of studies have investigated the CFA context. Olsson, Troye, and Howell (1999) evaluated the effects of the estimation method (ML vs. generalized least squares), model misspecification (defined as parametric model misspecifications—i.e., adding or removing paths), and sample size on the recovery of the underlying structure (which they called “the theoretical fit”) and the goodness of fit (which they called “the empirical fit”). Their results suggested better theoretical fit for ML, but at the cost of lower empirical fit. In addition, they found that misspecification exerted a large effect on both the theoretical and empirical fits. More recently, Ximénez (2006) conducted a simulation study on the recovery of weak factor loadings in CFA under varying conditions of estimation method (ML vs. ULS), sample size ($N = 100, 300$, and 500), loading size for the weak factor ($.25$, $.35$, or $.50$), model specification (correct vs. incorrect by altering the number of factors), and factor correlation (null vs. moderate). The results showed that the recovery of weak factor loadings improved when the factors were correlated and the models were correctly specified. For incorrectly specified models, the recovery was satisfactory when the misspecification implied overfactoring. However, in conditions of misspecification by underfactoring, the recovery was very poor, especially for models with orthogonal factors. In addition, the ULS method produced more convergent solutions and successfully recovered the weak factor loadings in some instances in which ML failed.

The Ximénez (2006) study extended previous research in several ways because it referred to CFA, used lower loading sizes to define the weak factor loadings, referred to models with orthogonal and correlated factors, and included misspecification conditions. However, more research is needed to continue examining these effects under more realistic conditions. For instance, previous studies in the context of CFA have only considered parametric model misspecification conditions (i.e., adding or deleting paths); none have examined the recovery of weak factor loadings for models that do not exactly hold in the population. Efforts to extend this work to cases in which models do not hold in the population are described below.

As many authors have noted within the factor analysis literature (e.g., MacCallum, 2003; Thurstone, 1930), studies have most often been based on a population correlation matrix that exactly satisfies a factor analysis model, whereas in practice it is unlikely that any factor analysis model will perfectly fit a population matrix. Therefore, it is necessary to examine more realistic population matrices and study the recovery of weak factor loadings in CFA when the model is moderately to highly misspecified in the population. How do ML and ULS behave and perform

when the model is not correct in the population? How is the recovery of weak factor loadings affected? Given that models are always wrong to some degree, the answers to these questions could be highly relevant and informative for researchers in practice.

Designing studies to address these questions requires the simulation of artificial data incorporating model error. There are different manners of introducing such error. Tucker, Koopman, and Linn (1969) suggested that population matrices could be simulated by including three kinds of factors: major common factors (or a small group of dominant latent variables), unique factors (or the traditional specific effect plus the error associated with each variable), and minor common factors (representing small sources of covariance among variables). They proposed that the population covariance matrix is made up of a particular structure plus additional elements of covariance representing the lack of fit. This method has been effectively used in several empirical studies (e.g., Hakstian, Rogers, & Cattell, 1982). Cudeck and Browne (1992) proposed another method for constructing a covariance matrix, in which the specified departure or lack of fit between the population matrix and the model is operationalized as an exact value of a discrepancy function. The present article uses the Cudeck and Browne method to examine whether the recovery of weak factor loadings in CFA is affected when the model is moderately to highly misspecified in the population. This method was chosen because it has the advantage that there is no need to designate the specific nature of the model error. That is, instead of introducing a particular type of model error (e.g., omitting minor factors, introducing nonlinear relationships, etc.), it potentially includes all types of possible errors.

Error is usually understood as arising from two distinct sources: sampling error and model error (MacCallum, Browne, & Cai, 2007). *Sampling error* refers to the lack of correspondence between the sample and the population from which it was drawn. *Model error* refers to the lack of fit of a model within a population and, as stated above, may arise from different sources. In this article, the term *model error* is reserved for the kind of error introduced by Cudeck and Browne (1992). Moreover, the term *structural error* will be used for a mismatch between the factorial structures of the true and estimated models (e.g., parametric model misspecification by adding or removing paths).

Two simulation studies were conducted. Both introduced model error by manipulating values of the discrepancy function and evaluated a number of sampling error conditions. The first study examined the recovery of weak factor loadings for correctly specified models (i.e., models without structural error). The second study examined the recovery of weak factor loadings for models incorrectly specified by altering the number of factors (i.e., models with structural error). Only misspecification by an underfactoring condition, which consisted of omitting one factor from the model, was considered. On the one hand, this choice was based on the results of previous research, which indicated that the recovery of weak factor loadings

in CFA was especially poor when the model was incorrectly specified by underfactoring, whereas misspecification by overfactoring did not affect recovery (Ximénez, 2006). On the other (as noted by MacCallum et al., 2007), this condition reflects a realistic condition of applied research because, in the attempt to obtain a parsimonious model that accounts for the relationships among the measured variables, researchers tend to use a small number of factors.

The article is organized as follows. First, theoretical aspects are presented, including the Cudeck and Browne (1992) method and the framework from which the hypotheses are derived. Second, the design and results of the two simulation studies are presented. Finally, the General Discussion summarizes the results and their practical implications.

THEORETICAL BACKGROUND

The CFA model (Jöreskog & Sörbom, 1981) can be given as follows:

$$\mathbf{x} = \Lambda \boldsymbol{\xi} + \boldsymbol{\delta}, \quad (1)$$

where $\mathbf{x}$ is a random vector of p observed variables, $\boldsymbol{\xi}$ is a random vector of q factors such that $q < p$, Λ is a $p \times q$ matrix of factor loadings, and $\boldsymbol{\delta}$ is a random vector of p measurement error variables. It is assumed that $E(\mathbf{x}) = E(\boldsymbol{\xi}) = E(\boldsymbol{\delta}) = 0$ and that $E(\boldsymbol{\xi}\boldsymbol{\xi}') = 0$.

From Equation 1, one can derive a model for Σ , the population covariance matrix for the observed variables $\mathbf{x}$:

$$\Sigma = \Lambda \Phi \Lambda' + \Theta_{\delta}, \quad (2)$$

where Φ is the $q \times q$ population covariance matrix of $\boldsymbol{\xi}$, and Θ_{δ} is the $p \times p$ population covariance matrix of $\boldsymbol{\delta}$. For convenience, it is usually assumed that $\Phi = \mathbf{I}$ and that Θ_{δ} is diagonal. Under this model, the parameters Λ , Φ , and Θ_{δ} have fixed, true values in the population. The model can be fit to a sample covariance matrix, $\mathbf{S}$, by using a method defined by a discrepancy function (e.g., ML or ULS) and estimating parameters so as to minimize the value of that discrepancy function. If Σ were available and the model of interest were fit to it, the resulting solution would yield a parameter vector $\boldsymbol{\gamma}$ and an implied $p \times p$ covariance matrix $\Sigma(\boldsymbol{\gamma})$.

The Cudeck and Browne Procedure

Cudeck and Browne (1992) developed a procedure for constructing a covariance matrix, Σ_0^* , with a specified minimum discrepancy function value in the population. Let γ_0 be a particular value within the admissible region Γ at which $\Sigma(\gamma_0)$ is positive definite. Let $\mathbf{E}$ be a symmetric matrix such that the sum,

$$\Sigma_0^* = \Sigma(\gamma_0) + \mathbf{E}, \quad (3)$$

is positive definite. Given a general discrepancy function of the form

$$F[\Sigma_0^*, \Sigma(\boldsymbol{\gamma})] = (1/2)tr\{\mathbf{W}^{-1}[\Sigma_0^* - \Sigma(\boldsymbol{\gamma})]^2\}, \quad (4)$$

where $\mathbf{W}$ is a fixed matrix that does not depend on $\mathbf{E}$. Here we consider two discrepancy functions: ULS, which is obtained when $\mathbf{W} = \mathbf{I}$ in Equation 4, and ML, where γ_{ML} is the minimizer of

$$M[\Sigma_0^*, \Sigma(\boldsymbol{\gamma})] = \ln|\Sigma(\boldsymbol{\gamma})| - \ln|\Sigma_0^*| + tr[\Sigma_0^* \Sigma(\boldsymbol{\gamma})^{-1}] - p. \quad (5)$$

If $\mathbf{W} = \Sigma(\gamma_{ML})$, the minimizer of Equation 4 is the same as the minimizer of Equation 5.

Cudeck and Browne (1992) stated that this problem can be addressed as follows: Given a particular value γ_0 for the parameter vector and a value $\delta > 0$ for the lack of fit, we seek a matrix $\mathbf{E}$ in Equation 3 such that

(A) the minimizer of $F[\Sigma_0^*, \Sigma(\boldsymbol{\gamma})]$ is the required value $\boldsymbol{\gamma} = \gamma_0$, and

(B) at the point γ_0 , the minimum function value will be one of the following:

$$\begin{aligned} M[\Sigma_0^*, \Sigma(\gamma_0)] &= \delta, & \text{if } \mathbf{W} &= \Sigma(\gamma_0), \\ F[\Sigma_0^*, \Sigma(\gamma_0)] &= \delta, & \text{otherwise,} \end{aligned}$$

where δ is a prespecified value (see Cudeck & Browne, 1992, pp. 359–361, for details of the algorithms used for computing $\mathbf{E}$).

In the present study, model discrepancy is operationalized by the following values in δ : 0, .10, .20, .30, and .40. The $\delta = 0$ value has been chosen to represent the condition in which the model holds exactly, and the $\delta > 0$ values have been chosen to represent conditions in which the model does not hold to different degrees (from models moderately to highly misspecified).

Derivation of Research Hypotheses

In this section, the mathematical approach proposed by MacCallum et al. (2001) is used to derive a series of hypotheses. Given that a sample covariance matrix $\mathbf{S}$ differs from the population covariance matrix Σ because of sampling error (Δ_{SE}), the consequent lack of fit attributable to Δ_{SE} can be expressed by defining Equation 2 for the sample factor solutions as

$$\mathbf{S} = \Lambda \Phi \Lambda' + \Theta_{\delta} + \Delta_{SE}. \quad (6)$$

In an ideal case, in which the variances and covariances match the corresponding population values, Δ_{SE} will be null. MacCallum et al. (2001) found that as the sample size increases, the sample variances and covariances will tend to approach their population values, thus reducing the impact of Δ_{SE} and causing the sample factor solutions to become more similar to the population solutions. They also noted that the magnitude of the elements in Θ_{δ} plays an important role. As these weights increase (or, equivalently, if the factor loadings of the measured variables are low, as in the present study), the values of the elements in Θ_{δ} will be high, and therefore their elements will receive more weight in Equation 6. These will then make a larger contribution to the structure of $\mathbf{S}$, causing the poorer recovery of the population factors. In such a case, if the elements of Φ also receive more weight (i.e., if factors are correlated), this effect may be attenuated.

When model error is present in the population, Equation 2 can be expressed as follows:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta_{\delta} + \Delta_{ME(P)}, \quad (7)$$

where $\Delta_{ME(P)}$ represents the lack of fit of the model in the population (notice that the $\Delta_{ME(P)}$ term is equivalent to the **E** term in the Cudeck and Browne procedure).

When model error in the population is explicitly represented, as in Equation 7, different methods yield different parameter estimates. Here, the ML and ULS estimation methods will be compared in order to examine the hypothesis, congruent with previous research in the context of EFA, that in situations with a moderate amount of model error, ULS will perform better than ML in recovering weak factors. MacCallum et al. (2001) suggested that the degree of correspondence between the nature of the error in the data and the assumptions about error for each method may account for the poorer performance of ML. Under ML, all error is assumed to be sampling error, and discrepancies in the residual correlation matrix are differentially weighted such that those discrepancies associated with larger correlations are more highly weighted. In contrast, under ULS, discrepancies are weighted equally. ML, then, attempts to fit larger correlations that are the result of model error rather than major common factors, and neglects the smaller correlations corresponding to the weak factor, thus failing to recover the weak factor loadings.

The sample covariance matrix can be expressed as in Equation 7:

$$S = \Lambda\Phi\Lambda' + \Theta_{\delta} + \Delta_{SE} + \Delta_{ME(S)}. \quad (8)$$

This expression includes terms representing the lack of fit of the factor model due to sampling error and model error. A comparison of Equations 7 and 8 shows that any difference between these solutions arises from the roles of $\Delta_{ME(P)}$ in the population and $\Delta_{SE} + \Delta_{ME(S)}$ in the sample. Sampling error (Δ_{SE}) only affects the solution obtained from a sample, not that obtained from the population. However, model error affects both the sample and population solutions. To the extent that $\Delta_{ME(P)}$ and $\Delta_{ME(S)}$ are similar, the population factors will be recovered more accurately in analysis of the sample data. As noted above, MacCallum et al. (2001) found that, regardless of sample size, as long as the model is correctly specified, model error does not influence the recovery of population factors. This is not surprising, because of the definition of the **E** term in the Cudeck and Browne procedure and its property A.

With respect to the impact of model error when the model includes structural error, given that there has been no previous research on this topic, the present study is eminently exploratory. However, given that misspecification by underfactoring is a source of model error that only affects the sample solutions, the $\Delta_{ME(P)}$ and $\Delta_{ME(S)}$ elements in Equations 7 and 8 will be less similar in this situation, and the population factors and their factor loadings are expected to be recovered less accurately in the analysis of sample data. The aim of the present study is to examine the magnitude of this effect.

In summary, the following hypotheses were investigated: (1) As sample size increases, sampling error will be reduced, and the sample solutions will be more stable and recover the population weak factor loadings more accurately; (2) the recovery of weak factor loadings will improve if the factors are correlated; (3) ULS is expected to perform better than ML in the recovery of weak factor loadings; and (4) as long as the model is correctly specified, the recovery of population weak factor loadings is not expected to be influenced by the presence of model error. However, the recovery of weak factor loadings is expected to be affected by the presence of model error when the model is misspecified.

SIMULATION STUDIES

Two simulation studies were conducted. The first explored the effects of estimation method, sample size, model discrepancy, and factor correlation on the recovery of weak factor loadings in the context of CFA for correctly specified models. The second study was conducted to examine whether the results found for the models specified with the known correct number of factors held when the model was misspecified in an underfactoring condition, which consisted of omitting one factor from the model. Therefore, this second study considered both model and structural error. The effects of the independent variables on the goodness of fit of the model and on the occurrence of nonconvergent solutions and Heywood cases were also examined in both studies.

The next section presents the procedure and methods of analysis, which were common to both studies. Afterward, a detailed description is provided of the results for each study.

General Procedure

The general approach used in both studies involved the following four steps:

1. Population factor structures (or generating models) were defined on the basis of one of the models used in Ximénez (2006), which included 12 measured normal variables and three factors, of which the third factor was relatively weak. This model was chosen because it showed the most important statistical and practical effects as compared with one- and two-factor models. Moreover, Briggs and MacCallum (2003) used a similar model in their study in the context of EFA. Each factor was defined by 4 observed variables, and both orthogonal and correlated factor conditions were simulated. The theoretical values of the parameters for each factorial structure are summarized in the upper panels of Figure 1. The weak factor had loadings of .30, to distinguish it from the major factors, which had loadings of .80 or more.

The population factor structures were used as the basis to generate the population covariance matrices, which were defined under the assumption that the factor model does not exactly hold in the population. The specified departure, or lack of fit between the population matrix and the model, was operationalized as an exact value of the

Study 1 **Study 1** **Study 2** **Study 2**

Figure 1. Theoretical and fitted models used in the simulation studies. The upper plots in all panels represent the theoretical models used in Studies 1 and 2. In Study 1, the fitted models are those in the upper panels (i.e., the correct models). However, in Study 2, the fitted models are those in the lower panels (i.e., the models were incorrectly specified by omitting one factor). Thus, the weak factor has been contaminated by including some indicators that theoretically belong to another factor).

discrepancy function. A FORTRAN program was used to compute the population covariance matrices for each factorial structure, discrepancy value, sample size, and estimation method (ML vs. ULS) following the Cudeck and Browne procedure.

2. The population covariance matrices were used as the basis for simulating the sample covariance matrices. One thousand sample covariance matrices were simulated with the PRELIS 2 program of Jöreskog and Sörbom (1996b) for each model.

3. A CFA was conducted on each simulated sample covariance matrix using ML and ULS estimation. The parameter estimates were computed with the LISREL 8.80 program of Jöreskog and Sörbom (1996a).

4. The sample factor solutions were evaluated to determine how the recovery of weak factor loadings was affected by the independent variables of the study.

The independent variables for both studies were sample size, factor correlation, model discrepancy, and estimation method. A detailed description of the levels of these variables follows. The smallest sample size (N) chosen was 100, because it is dangerous to use ML CFA with sample sizes of less than 100, particularly for models with relatively low factor loadings (Boomsma, 1982). To approximate medium and relatively large sample sizes, 300 and 500 observations were used. Two levels of factor correlation were chosen: null = 0 and moderate = .50. As stated above, model discrepancy was introduced using the procedure developed by Cudeck and Browne and operationalized by the following values in δ : 0, .10, .20, .30, and .40 (where the $\delta = 0$ value means that the model holds exactly, and the $\delta > 0$ values that the model does not hold

to different degrees, from models moderately to highly misspecified). To facilitate the interpretation of these values for researchers in the lab, the $\delta > 0$ values, translated into RMSEA, correspond to .022, .043, .065, and .086, respectively. Finally, the ML and ULS estimation methods were considered. The dependent variables were the recovery of weak factor loadings, goodness of fit, and the occurrence of nonconvergent solutions and Heywood cases. The variables in the overall design are summarized in Table 1.

Analyses of Output

Nonconvergent solutions (NCONVER) were deleted to study the effects of the independent variables on the recovery of the weak factor loadings. The operational definition employed was that of the LISREL program: failure to reach convergence after 250 iterations (see Jöreskog, 1967, p. 460). Moreover, Heywood cases were detected in each of the cells of the design, but for analysis purposes were not deleted. The nonconvergent solutions were analyzed separately to study the effect of the independent variables on the occurrence of nonconvergent solutions and Heywood cases. Two qualitative variables were created. For NCONVER, nonconvergent solutions were coded "1," whereas convergent solutions were coded "0." For Heywood cases (HEYWOOD), solutions with Heywood cases were coded "1," whereas solutions without Heywood cases were coded "0." Loglinear-logit models were fitted to the data using ML estimation. The proportion of weighted variation explained by each model was calculated, in addition to the usual likelihood ratio chi-square statistic. The D measure proposed by McFadden

Table 1
Variables Considered in the Monte Carlo Study

Code	Variable	Levels
Independent Variables		
M	Method	ML (maximum likelihood) ULS (unweighted least squares)
N	Sample size	100 300 500
D	Model discrepancy, δ	.00 .10 .20 .30 .40
C	Correlation between factors	.00 .50
Dependent Variables		
ϕ	Coefficient of congruence	
RMSD	Root-mean squared deviation	
RMSEA	Root-mean squared error of approximation	
NCONVER	Nonconvergent solutions	0: no 1: yes
HEYWOOD	Heywood cases	0: no 1: yes

Note—A $2 \times 3 \times 5 \times 2$ design was used for both Studies 1 and 2.

(1974) was used to measure the proportion of variance explained by each model.

Recovery of the weak factor loadings was assessed by inspection of the correspondence between the theoretical and estimated loadings for the weak factor only. Two measures of correspondence were used. The first was the coefficient of congruence (Tucker, 1951):

$$\phi_k = \frac{\sum_{i=1}^p \lambda_{ik(t)} \lambda_{ik(e)}}{\sqrt{\left(\sum_{i=1}^p \lambda_{ik(t)}^2 \right) \left(\sum_{i=1}^p \lambda_{ik(e)}^2 \right)}}, \quad (9)$$

where p is the number of variables that define the factor k , $\lambda_{ik(t)}$ is the theoretic loading for the observed variable i of the factor k , and $\lambda_{ik(e)}$ is the corresponding loading obtained from the simulation data. The same interpretation guidelines were adopted as in MacCallum et al. (2001): Values of ϕ above .98 indicate excellent recovery; from .92 to .98, good recovery; from .82 to .92, borderline recovery; from .68 to .82, poor recovery; and below .68, terrible recovery.

A second measure of correspondence, the root-mean square deviation (RMSD; Levine, 1977) was also calculated for the weak factor only:

$$\text{RMSD}_k = \sqrt{\sum_{i=1}^p (\lambda_{ik(t)} - \lambda_{ik(e)})^2 / p}. \quad (10)$$

RMSD reaches a minimum of 0 for a perfect pattern-magnitude match and a maximum of 2 when all loadings are equal to unity but of opposite signs. In practice, most studies consider that RMSD values below .20 are indicative of satisfactory recovery.

The two measures of correspondence were used in both Studies 1 and 2. However, notice that in Study 2, given the

structural error, the fitted models were different from the generating or theoretical models, although this difference was small (see Figure 1). As stated above, the simulated situation reflects a realistic situation in applied research, when a factor has been contaminated by including some indicators that theoretically belong to another factor. Thus, the correspondence measures assessed how well the weak factor loadings were recovered in the presence of two contaminating indicators that theoretically belonged to another factor.

A simple metamodel was used to analyze the results, which included only the main and the double interaction effects of each independent variable on the dependent variable. Following Skrondal (2000, pp. 145–146), interactions of three factors or higher were discarded because of the tenet of parsimony, because their interpretation is strenuous, and because discarding higher order interactions may improve precision. The following model was tested:

$$\text{RWFL} = \mu + M + N + D + C + M \times N + M \times D + M \times C + N \times D + N \times C + D \times C, \quad (11)$$

where RWFL = recovery of weak factor loadings (ϕ and RMSD measures), M = method (ML vs. ULS), N = sample size (100, 300, or 500), D = model discrepancy value ($\delta = 0, .10, .20, .30$, or $.40$), and C = correlation between factors (0 or $.50$).

A four-way ANOVA was conducted to test the effects included in the metamodel. All of the effects were viewed as independent. Since a large sample size ($N = 60,000$) can cause even negligible effects to be statistically significant, the explained variance associated with each of the effects was also calculated, measured by the η^2 statistic. The interpretation guidelines suggested by Cohen (1988) were adopted: η^2 values from .05 to .09 indicate a small effect;

Table 2
Proportions of Nonconvergent Solutions and Heywood Cases Across
the Independent Variables of Simulation Studies 1 and 2

D	C = 0						C = .50					
	N = 100		N = 300		N = 500		N = 100		N = 300		N = 500	
	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS
Study 1												
NCONVER												
.00	.41	.34	.24	.18	.11	.09	.10	.05	.01	.00	.00	.00
.10	.41	.36	.24	.24	.13	.17	.10	.05	.01	.01	.00	.00
.20	.42	.35	.26	.26	.16	.20	.11	.06	.01	.01	.00	.01
.30	.42	.38	.27	.35	.18	.32	.11	.07	.01	.03	.00	.01
.40	.42	.42	.29	.43	.22	.40	.11	.06	.02	.02	.00	.02
HEYWOOD												
.00	.48	.42	.27	.21	.13	.10	.31	.24	.06	.05	.01	.01
.10	.48	.48	.28	.29	.16	.18	.33	.26	.07	.07	.02	.02
.20	.48	.44	.29	.30	.19	.22	.33	.26	.07	.08	.02	.04
.30	.49	.47	.30	.39	.21	.33	.32	.28	.08	.11	.02	.06
.40	.50	.49	.33	.46	.25	.42	.33	.29	.09	.13	.03	.08
Study 2												
NCONVER												
.00	.33	.08	.33	.09	.34	.10	.00	.00	.00	.00	.00	.00
.10	.40	.20	.48	.36	.57	.46	.00	.00	.00	.00	.00	.00
.20	.46	.29	.59	.48	.69	.60	.00	.00	.00	.00	.00	.00
.30	.49	.31	.65	.54	.76	.66	.00	.00	.00	.00	.00	.00
.40	.44	.36	.25	.58	.17	.69	.00	.00	.00	.00	.00	.00
HEYWOOD												
.00	.71	.47	.74	.50	.74	.51	.02	.00	.00	.00	.00	.00
.10	.77	.75	.88	.87	.92	.91	.04	.00	.00	.00	.00	.00
.20	.84	.82	.92	.95	.97	.98	.05	.00	.00	.00	.00	.00
.30	.87	.86	.96	.96	.99	.98	.07	.00	.01	.00	.00	.00
.40	.90	.88	.97	.98	.98	.99	.09	.01	.01	.00	.00	.00

Note—C, correlation between factors; N, sample size; D, model discrepancy value; ML, maximum likelihood; ULS, unweighted least squares; NCONVER, nonconvergent solutions; HEYWOOD, Heywood cases.

from .10 to .20, a medium effect; and above .20, a large effect. Multiple comparisons were also conducted for the effects that were shown to be statistically and practically significant.

The goodness of fit of the model was measured by the root-mean squared error of approximation (RMSEA) index of Steiger (1990). RMSEA was chosen because it showed good performance in a simulation study by Hu and Bentler (1999) and because it displays an interpretable scale for determining the degree of fit. Browne and Cudeck (1993) suggested that values of RMSEA below .05 indicate close fit; from .05 to .08, fair fit; from .08 to .10, mediocre fit; and above .10, unacceptable fit. In addition, RMSEA is sensitive to model misspecification (Fan & Sivo, 2005). The same metamodel as in Equation 11 was used to test the effects of the independent variables on the RMSEA index by a four-way ANOVA.

Results of Simulation Study 1

Nonconvergence and Heywood cases. Of the 60,000 solutions, 9,694 (16.2%) were nonconvergent, and 14,150 (23.6%) presented Heywood cases. The proportions of nonconvergent solutions and Heywood cases that occurred in obtaining 1,000 good solutions per cell are summarized in the upper section of Table 2. The results of the loglinear-logit analyses are summarized in the upper section of Table 3.

The two-way interaction models provided good explanations of the data, accounting for more than 99% of the weighted variation to be explained for both nonconvergent solutions and Heywood cases. Examination of the parameter estimates and the chi-square values for NCONVER and HEYWOOD showed that the proportion of nonconvergent and improper solutions decreased when the factors were correlated, the sample size was increased, and the model discrepancy was smaller. The $N \times D$, $M \times N$, and $M \times D$ interaction effects were of considerable size. Analyses showed that the effect of model discrepancy was most pronounced for the smallest sample size ($N = 100$). In addition, for small and medium sample sizes ($N = 100$ and 300), there were fewer nonconvergent and improper solutions with the ULS estimation method. Finally, the proportion of nonconvergent solutions increased for ULS solutions as model discrepancy also increased.

Recovery of weak factor loadings. The upper section of Table 4 shows the summary statistics for the measures of recovery of weak factor loadings (ϕ and RMSD) for all of the main effects. The upper section of Table 5 presents the results of the ANOVA for the RMSD measure. The ANOVA results for the congruence measure, ϕ , are not included for brevity, because they are very similar to the RMSD results.

As shown in Table 5, all of the main effects and nearly all of the double interactions were statistically significant.

Table 3
Effect of Independent Variables on the
Nonconvergent Solutions and Heywood Cases

		NCONVER		HEYWOOD	
	<i>df</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
Study 1					
M	1	61.327	<.001	66.144	<.001
N	2	1,493.237	<.001	4,017.668	<.001
D	4	304.472	<.001	342.310	<.001
C	1	4,317.711	<.001	8,177.530	<.001
M × N	2	138.776	<.001	122.705	<.001
M × D	4	72.943	<.001	73.947	<.001
M × C	1	0.003	.999	0.001	.978
N × D	8	159.373	<.001	218.958	<.001
N × C	2	0.001	.999	0.001	.999
D × C	4	0.006	.999	0.001	.999
P			.993		.994
Study 2					
M	1	137.196	<.001	55.574	<.001
N	2	449.887	<.001	19.204	<.001
D	4	1,452.533	<.001	1,222.879	<.001
C	1	7,457.331	<.001	8,332.117	<.001
M × N	2	196.582	<.001	92.556	<.001
M × D	4	907.503	<.001	1,462.315	<.001
M × C	1	0.001	.971	0.002	.989
N × D	8	161.534	<.001	151.636	<.001
N × C	2	0.001	.999	0.001	.999
D × C	4	0.001	.999	0.001	.999
P			.990		.995

Note—NCONVER, nonconvergent solutions; HEYWOOD, Heywood cases; M, method; N, sample size; D, model discrepancy value; C, correlation between factors; P, proportion of weighted variation explained by each model.

The largest effects found were due to the sample size ($\eta^2 = .20$) and factor correlation ($\eta^2 = .08$) main effects. The recovery of weak factor loadings improved as the sample size increased. As can be seen from Table 4, the average values of ϕ and RMSD for the smallest sample size ($N = 100$) were indicative of terrible recovery, and those for the medium and large sample sizes ($N = 300$ and 500) respectively indicated borderline and good recovery. The presence of factor correlation significantly improved the weak factor loadings' recovery: The average values of the correspondence measures for orthogonal factors were indicative of poor recovery, and those for the correlated factors of satisfactory recovery. The $N \times C$ interaction produced a statistically significant effect, but its effect size was very small ($\eta^2 = .002$). Figure 2A illustrates the absence of this interaction. As shown, the recovery of weak factor loadings was satisfactory in all the sample sizes when the factors were correlated. However, the recovery worsened if the factors were orthogonal, and was especially poor for the smallest sample size.

Estimation method also produced a statistically significant, though very small, effect ($\eta^2 = .013$). Overall, the mean values for both ϕ and RMSD indicated that the recovery of weak factor loadings with the ULS estimation method was slightly better than with the ML method (see the upper section of Table 4). The scatterplots in the first and second rows of Figure 3 illustrate this difference in more detail. These plots show the RMSD coefficient for the weak factor loadings from the ML and ULS solu-

tions for the convergent cases under varying conditions of model discrepancy and correlation (to conserve space, the plots with the sample size conditions are not included, but are available from the author on request). As shown, when the factors were correlated (see the plots from Figures 3F to 3J), the majority of the points were concentrated in the lower left corner, representing replications in which both ML and ULS adequately recovered the weak factor loadings. In many other instances, however, ULS recovered the weak factor loadings satisfactorily, but ML did not. This was reflected by the points in the plot above 0.20 on the horizontal axis and below 0.20 on the vertical axis (these corresponded to 10% of cases, which were models with $N = 100$ that were not associated with the occurrence of Heywood cases). There were also cases in which both methods obtained high values in RMSD (these corresponded to models with $N = 100$). When the factors were orthogonal (see the plots from Figures 3A to 3E), the recovery of weak factor loadings was poorer. Both the ML

Table 4
Summary Statistics on Dependent Variables for
Main Effects in Simulation Studies 1 and 2

	Congruence (ϕ)		RMSD		RMSEA	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Study 1						
Overall	.8331	.3297	0.1611	0.1467	.0970	.0642
M						
ML	.8012	.3318	0.1630	0.1480	.0580	.0402
ULS	.8651	.3276	0.1592	0.1454	.1365	.0596
N						
100	.6687	.4792	0.2599	0.1940	.0967	.0649
300	.8693	.2640	0.1411	0.1120	.0972	.0637
500	.9222	.1593	0.1060	0.0846	.0974	.0642
D						
.00	.8390	.3307	0.1549	0.1400	.0455	.0499
.10	.8384	.3219	0.1576	0.1446	.0795	.0498
.20	.8378	.3181	0.1603	0.1465	.1009	.0490
.30	.8262	.3377	0.1651	0.1492	.1055	.0379
.40	.8231	.3403	0.1686	0.1534	.1384	.0372
C						
.00	.7994	.3048	0.2023	0.1676	.1191	.0703
.50	.8611	.3465	0.1270	0.1163	.0787	.0520
Study 2						
Overall	.3341	.7111	0.2083	0.1729	.1955	.0366
M						
ML	.3194	.7103	0.2278	0.2249	.1670	.0182
ULS	.3488	.7116	0.1888	0.0919	.2241	.0268
N						
100	.3426	.6428	0.2247	0.2124	.1929	.0382
300	.3386	.7261	0.2032	0.1577	.1966	.0359
500	.3210	.7592	0.1970	0.1388	.1970	.0354
D						
.00	.6526	.5085	0.1571	0.0821	.1870	.0307
.10	.3836	.6808	0.1910	0.0836	.1928	.0320
.20	.3428	.7124	0.1927	0.0842	.1931	.0295
.30	.3072	.7257	0.1950	0.0850	.1962	.0527
.40	.1842	.7881	0.2757	0.3401	.2084	.0287
C						
.00	.2264	.6016	0.2968	0.2055	.2007	.0364
.50	.8947	.1449	0.1198	0.0432	.1903	.0360

Note—M, method; N, sample size; D, model discrepancy value; C, correlation between factors.

Table 5
ANOVA Results for the Dependent Variables in Simulation Studies 1 and 2

		RMSD			RMSEA		
	<i>df</i>	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Study 1							
M	1	525.22	<.001	.013	101,571.44	<.001	.669
N	2	6,144.16	<.001	.198	114.65	<.001	.005
D	4	23.61	<.001	.002	9,540.18	<.001	.432
C	1	4,261.33	<.001	.078	25,922.56	<.001	.340
M × N	2	12.81	<.001	.001	3.75	.024	.000
M × D	4	9.00	<.001	.001	818.96	<.001	.061
M × C	1	37.95	<.001	.001	13,310.73	<.001	.209
N × D	8	0.48	.870	—	43.46	<.001	.007
N × C	2	46.17	<.001	.002	1.57	.208	—
D × C	4	0.35	.843	—	5,977.47	<.001	.322
Error	50276	(.016)			(.001)		
Total	50306						
				.295			.803
Study 2							
M	1	3,715.74	<.001	.073	154,346.50	<.001	.766
N	2	591.62	<.001	.024	549.32	<.001	.023
D	4	3,573.35	<.001	.232	2,458.51	<.001	.172
C	1	123,070.62	<.001	.723	8,679.14	<.001	.155
M × N	2	30.64	<.001	.001	7.07	<.001	.000
M × D	4	1,081.06	<.001	.084	1,200.24	<.001	.092
M × C	1	321.53	<.001	.007	0.29	.593	—
N × D	8	7.20	<.001	.001	12.85	<.001	.002
N × C	2	335.75	<.001	.015	45.52	<.001	.002
D × C	4	3,274.38	<.001	.217	4,268.09	<.001	.266
Error	47208	(.003)			(.0003)		
Total	47238						
				.780			.825

Note—Values in parentheses represent mean squared errors. RMSD, root-mean squared deviation; RMSEA, root-mean squared error of approximation; M, method; N, sample size; D, model discrepancy value; C, correlation between factors.

and ULS solutions showed similar results in the majority of these cases, but there were still some cases in which ML failed yet ULS succeeded (these corresponded to 12% of cases). Overall, these plots also showed that recovery of weak factor loadings worsened when factors were defined as orthogonal.

Finally, even though the main effect of model discrepancy was statistically significant, its effect size was very small ($\eta^2 = .002$). Overall, the mean values for both ϕ and RMSD (see the upper section of Table 4) indicated that in those conditions in which the structure did not exactly hold ($\delta > 0$), the recovery of weak factor loadings was essentially equal to that in which the structure held ($\delta = 0$). Thus, the results showed no appreciable influence of model discrepancy on the correspondence between the sample and population weak factor loadings. This finding was consistent across the levels of the remaining design features considered (estimation method, sample size, and factor correlation).

Goodness of fit. The summary statistics on RMSEA for all of the main effects and the ANOVA results appear in the upper right sections of Tables 4 and 5. As shown in Table 5, the largest effects were attributable to the estimation method ($\eta^2 = .67$), model discrepancy ($\eta^2 = .43$), and factor correlation ($\eta^2 = .34$) main effects. The D $\times$ C, M $\times$ C, and M $\times$ D interactions also produced effects ($\eta^2 = .32$, .21, and .06, respectively). The average values of RMSEA for models that held exactly in the population ($\delta = 0$) were

indicative of close fit, and those for models that did not hold exactly ($\delta > 0$) were indicative of mediocre or unacceptable fit. Therefore, the RMSEA measure was sensitive to model error. This effect held when factors were correlated. However, when factors were orthogonal, the fit was poor, even for models that held in the population (see the plot for the D $\times$ C interaction in Figure 2B). In addition, the mean RMSEA values were smaller for ML than for ULS (see Table 4). This effect was moderated by the effects of correlation and model discrepancy. The M $\times$ C interaction is represented in Figure 2C. The results indicated that the difference between the methods was not strong if the factors were correlated but was if they were orthogonal. The M $\times$ D interaction (represented in Figure 2D) indicated that when the model held exactly, only ML showed a close fit. However, when it did not hold, the average values of RMSEA were indicative of a mediocre fit for ML solutions and an unacceptable fit for the ULS solutions.

The ANOVA analyses for RMSD and RMSEA were repeated after eliminating the Heywood cases. The results are not included for brevity, because they replicated the previous ones. Thus, we may conclude that the presence of Heywood cases did not considerably influence the effects discussed above.

Results of Simulation Study 2

Nonconvergence and Heywood cases. Of the 60,000 solutions, 12,762 (21.3%) were nonconvergent, and 23,941

Figure 2. Graphical representation of the strongest double interaction effects found for the dependent variables of Studies 1 and 2.

Figure 3. Scatterplots for the RMSD measure across estimation methods, model discrepancies, and correlations in Studies 1 and 2.

(39.9%) presented Heywood cases. The proportions of nonconvergent solutions and Heywood cases that occurred in obtaining 1,000 good solutions per cell and the results of the loglinear–logit analyses are summarized in the lower sections of Tables 2 and 3, respectively. The two-way interaction models provided good explanations of the data, accounting for at least 99% of the weighted variation to be explained for NCONVER and HEYWOOD. Examination of the parameter estimates and the chi-square values showed that the proportion of nonconvergent and improper solutions decreased when the factors were correlated, the model discrepancy was reduced, and the sample size increased. Furthermore, there were more nonconvergent and improper solutions with the ML estimation method. The $M \times D$, $M \times N$, and $N \times D$ interaction effects were of considerable size. Analyses showed that the proportion of nonconvergent solutions increased for ML solutions as the model discrepancy also increased. In addition, for small and medium sample sizes ($N = 100$ and 300), there were fewer nonconvergent and improper solutions with the ULS estimation method and with lower model discrepancy values. Finally, the greatest proportion of Heywood cases occurred for ML solutions when the factors were orthogonal, whereas for correlated factors nearly all of the solutions were convergent and did not present Heywood cases.

Recovery of weak factor loadings. The lower sections of Tables 4 and 5 present the summary statistics for the measures of recovery of weak factor loadings and the ANOVA results for the RMSD measure. Recall that what was assessed in this case was the recovery of the weak factor population loadings when the weak factor was contaminated by including some indicators that theoretically belonged to another factor. As before, the ANOVA results for ϕ are not included because they were very similar to the RMSD results.

As shown in Table 5, all of the main effects and double interactions were statistically significant. The largest effects found were attributable to the main effects of factor correlation ($\eta^2 = .72$) and model discrepancy ($\eta^2 = .23$) and to the $D \times C$ interaction ($\eta^2 = .22$). The recovery of weak factor loadings for models incorrectly specified by underfactoring was much improved when the factors were correlated. As can be seen from Table 4, the average values of ϕ and RMSD for models with orthogonal factors were indicative of terrible recovery, whereas those for correlated factors were indicative of satisfactory recovery. As expected, the presence of model error for incorrectly specified models affected the recovery of weak factor loadings. The average values of the correspondence measures for models that held in the population ($\delta = 0$) were indicative of very poor recovery (as explained below, this was associated with the occurrence of Heywood cases); however, those for models that did not hold ($\delta > 0$) were indicative of terrible recovery. Figure 2E illustrates the $D \times C$ interaction. As shown, the recovery was satisfactory across all the discrepancy values when the factors were correlated. However, it worsened if the factors were orthogonal, and was especially poor for the most extreme value of model discrepancy ($\delta = .40$). Therefore, when the model included structural error, the results showed the

influence of model discrepancy on the correspondence between the sample and population weak factor loadings.

The estimation method and the $M \times D$ interaction also produced statistically significant, though small, effects ($\eta^2 = .07$ and $.08$, respectively). Overall, the recovery of weak factor loadings with the ULS estimation method was slightly better than with the ML method. Figure 2F illustrates the $M \times D$ interaction. As shown, recovery was slightly better for ULS than for ML solutions, and no differences in the ULS solutions were attributable to model discrepancy. However, for ML solutions, recovery was especially poor for the most extreme case of model discrepancy ($\delta = .40$). This was also attributable to the occurrence of Heywood cases. The scatterplots in the third and fourth rows of Figure 3 illustrate the differences between the estimation methods in more detail. As shown, when the factors were correlated (see the plots in Figures 3P to 3T), the majority of the points were concentrated in the lower left corner, representing replications in which both ML and ULS adequately recovered the weak factor loadings. In many other instances, ULS recovered the weak factor loadings satisfactorily and ML did not. It should be noted that in no case did ML appreciably outperform ULS in the recovery of weak factor loadings. When the factors were orthogonal (see Figures 3K to 3O), the plots showed a very different pattern. When the model held exactly in the population, the pattern of the differences was similar to that explained for correlated factors. However, as the model discrepancy increased, recovery became poorer. This was particularly clear for the most extreme case ($\delta = .40$), in which the recovery was especially poor for ML solutions.

Finally, the main effect of sample size and the $N \times C$ interaction were statistically significant, but their effect sizes were very small ($\eta^2 = .02$ for both). The $N \times C$ interaction is represented in Figure 2G. As shown, recovery was satisfactory across all sample size levels for correlated factors. However, for orthogonal factors, recovery was poor for all sample sizes, even the largest ($N = 500$).

Table 6 presents the summary statistics and ANOVA results for the RMSD measure after eliminating the Heywood cases. Again, the ANOVA results for ϕ are not included because they are very similar to the RMSD results. The results showed that some of the effects were associated with the occurrence of Heywood cases. For instance, eliminating Heywood cases improved the recovery for low values of model discrepancy. That is, recovery was poor for models that held in the population because of the presence of Heywood cases. However, eliminating Heywood cases did not improve the recovery for models that did not hold, in which recovery was poor especially for the largest discrepancy values ($\delta > .20$). In addition, after eliminating the Heywood cases, the $M \times D$ interaction effect was much smaller (η^2 went from $.08$ to $.01$), indicating that ULS performed slightly better than ML. Recovery was especially poor for the most extreme case of model discrepancy ($\delta = .40$) in both the ML and ULS solutions.

Goodness of fit. The summary statistics on RMSEA for all of the main effects and the ANOVA results appear in the lower right sections of Tables 4 and 5. As in Study 1,

Table 6
Results of Study 2 After Eliminating the Heywood Cases

		Summary Statistics					
		Congruence (ϕ)		RMSD		RMSEA	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Overall		.6936	.5276	0.1599	0.1117	.1952	.0389
M							
ML		.6102	.6141	0.1862	0.1336	.1650	.0196
ULS		.7812	.3996	0.1323	0.0731	.2269	.0271
N							
100		.6174	.4999	0.1792	0.1073	.1930	.0407
300		.7208	.5269	0.1533	0.1114	.1960	.0381
500		.7450	.5470	0.1466	0.1138	.1967	.0376
D							
.00		.8333	.2887	0.1312	0.0610	.1849	.0318
.10		.8252	.3122	0.1322	0.0625	.1919	.0342
.20		.7838	.3821	0.1387	0.0692	.1925	.0495
.30		.6546	.5319	0.1573	0.0803	.1980	.0360
.40		.4434	.7631	0.2251	0.1816	.2072	.0338
C							
.00		.2587	.6248	0.2953	0.1360	.2009	.0371
.50		.8972	.1382	0.1193	0.0427	.1906	.0360
		ANOVA Results					
		RMSD			RMSEA		
		<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
M	1	1,237.90	<.001	.033	89,677.73	<.001	.713
N	2	372.66	<.001	.020	481.26	<.001	.026
D	4	4,109.22	<.001	.313	3,626.20	<.001	.287
C	1	52,093.10	<.001	.591	6,850.55	<.001	.160
M $\times$ N	2	35.74	<.001	.002	36.95	<.001	.002
M $\times$ D	4	66.12	<.001	.007	188.58	<.001	.021
M $\times$ C	1	48.06	<.001	.001	1,900.33	<.001	.050
N $\times$ D	8	5.02	<.001	.001	8.62	<.001	.002
N $\times$ C	2	138.30	<.001	.008	96.97	<.001	.005
D $\times$ C	4	3,822.50	<.001	.298	2,168.05	<.001	.194
Error	36029	(.002)			(.0002)		
Total	36059			.823			.856

Note—Values in parentheses represent mean squared errors. RMSD, root-mean squared deviation; RMSEA, root-mean squared error of approximation; M, method; N, sample size; D, model discrepancy value; C, correlation between factors.

the largest effects were attributable to the main effects of estimation method ($\eta^2 = .77$), model discrepancy ($\eta^2 = .17$), and factor correlation ($\eta^2 = .16$), and to the D $\times$ C and M $\times$ D interactions ($\eta^2 = .27$ and $.09$, respectively). When the models included structural error, both the ML and ULS solutions displayed unacceptable fit. This result even occurred for models that held exactly in the population (see Figure 2I). For ML solutions, the empirical fit became poorer as model discrepancy increased. However, RMSEA values were indicative of an unacceptable fit across all discrepancy values for both ML and ULS solutions. This held for models with both orthogonal and correlated factors (see Figure 2H). That is, factor correlation did not improve the empirical fit.

SUMMARY AND GENERAL DISCUSSION

This article has focused on the recovery of weak factor loadings in CFA in the presence of model error. Model error was introduced using a procedure developed by Cudeck and Browne (1992) that allows the user to specify a

covariance structure with a specified discrepancy in the population. The results of two simulation studies examining the recovery of weak factor loadings in CFA under varying conditions of estimation method (ML vs. ULS), sample size, model discrepancy, and factor correlation were presented. The first study examined the recovery for models specified with the known correct number of factors (i.e., for models without structural error), and the second examined the recovery for models that included structural error (i.e., were incorrectly specified by underfactoring). The effects of the same variables on goodness of fit and the occurrence of nonconvergent solutions and Heywood cases were also examined.

The present work extends previous research by examining the recovery of weak factor loadings in CFA in two ways. First, the impact of model error on the recovery of weak factor loadings in the context of CFA had not previously been studied, and this study specifically addressed this issue. The present study represents a realistic condition for researchers in the lab because models, in their attempt to provide a representation of psychological phe-

nomena, are wrong to some degree and cannot be made exactly correct. At best, they provide approximations. Second, previous research examining the role of model error in the EFA context has only considered models without structural error (i.e., correctly specified models), and this study considers the situation in which a model includes structural error (i.e., the model is incorrectly specified by underfactoring). This also represents a realistic condition of applied research because researchers, in their attempt to obtain a parsimonious model that accounts for the relationships among the measured variables, tend to use a small number of factors.

The results of Simulation Study 1, focusing on models without structural error, found several significant effects that supported our hypotheses: (1) As sample size increased, sampling error was reduced and the sample solutions were more stable and recovered the population weak factor loadings more accurately (with a sample size of 300 or more observations being enough for adequate recovery); (2) the recovery of weak factor loadings under model error improved when the factors were correlated, producing satisfactory recovery even for small sample sizes (e.g., $N = 100$), whereas for models with orthogonal factors, recovery was much poorer and required a larger sample size; (3) under model error, ULS performed slightly better than ML and recovered the weak factor loadings in some instances in which ML failed; and (4) as long as the model was correctly specified, the recovery of population weak factor loadings was unaffected by the presence of model error. However, the number of nonconvergent solutions and Heywood cases did increase especially when the factors were orthogonal, the sample size was small, and the discrepancy value was large. These results are similar to those obtained by Anderson and Gerbing (1984) for ML solutions and models without structural error. The results in terms of empirical fit indicated that RMSEA was sensitive to model error. If the factors were correlated, the RMSEA values were indicative of a close fit when the model held in a population, and of a mediocre or unacceptable fit when the model did not. However, when the factors were orthogonal, the fit across all discrepancy values was poor and the empirical fit was better for ML than for ULS solutions. These findings are consistent with those of la Du and Tanaka (1989) and Ximénez (2006).

The results of Simulation Study 2 indicated that the effects found for correctly specified models did not generalize to models including structural error. The presence of model error produced a large effect on the recovery of the weak factor population loadings when the weak factor was contaminated with some indicators that theoretically belonged to another factor. As in Study 1, if the factors were orthogonal, the recovery was poor for models that held in the population and was much poorer for models that did not hold. The poor recovery for models that held was found to be associated with the occurrence of Heywood cases. However, if the factors were correlated, the recovery was adequate across all model discrepancy values, and the results were similar to those found in Study 1 for models without structural error. In addition, the effect of sample size was not as important when the model included

structural error, and as in Study 1, ULS produced slightly better results than did ML. The results in terms of nonconvergent solutions and Heywood cases indicated that when the model included structural error, a much larger number of improper solutions occurred, especially if the factors were orthogonal, the model discrepancy was large, and the sample size was small. Finally, the results in terms of goodness of fit indicated that the fit was much poorer than in Study 1. This is to be expected because Study 2 considered two sources of model error. The results of Study 2 are consistent with previous work undertaken in the context of EFA and CFA without model error (e.g., Fava & Velicer, 1996; Ximénez, 2006), which indicates that major distortion in the loading patterns occurs when the estimated model has fewer factors than the population model. Given that there is no previous research concerning this topic under conditions of model error, one aim of the present research was to examine the magnitude of the model error effect. The results demonstrated that model error produced a large effect on the recovery of weak factor loadings for models including structural error, indicating that when the factors were independent, the recovery worsened as model error increased and as sample size decreased. However, when the factors were correlated, the recovery was satisfactory regardless of the presence of model error.

At one level, the results of the present study give insights into the recovery of weak factor loadings in CFA when the model does not hold in the population, which is a more realistic condition for the factor analysis applications. At another level, some results have implications for the practical use of CFA with factorial structures that include weak factor loadings and incorporate model error, a situation that is always present to some degree in practice. These issues are related to aspects of a research project's design. First, the present study demonstrates how important it is to define factors as correlated for the adequate recovery of weak factor loadings. It should be noted that in the context of CFA, a majority of studies have allowed correlations among factors. However, some studies use orthogonal factors because an EFA with varimax rotation has been conducted first, and on the basis of these results a CFA is conducted, thus defining factors as orthogonal. Moreover, as a reviewer pointed out, in practice one cannot choose the amount of correlation between factors, or the estimate of correlation may be null. In such cases, researchers must be aware that the recovery of weak factor loadings and the empirical fit will be much poorer than if the factors were correlated. Second, this research has also demonstrated that, when the data come from a population structure in which all of the factors are not equally strong and there is a moderate amount of model error, ML fails to recover the weak factor loadings in some instances in which ULS succeeds. Therefore, researchers performing a CFA with factorial structures including weak factor loadings should favor the use of ULS estimation, or should at least compare the ML and ULS solutions. This represents important advice for applied researchers who, in many cases, erroneously believe that under multivariate normality ML is the only method available for estimating a CFA model. Third, the sample size must be much larger than

typically recommended (e.g., a 20:1 ratio of subjects to variables) when the model includes weak factor loadings and the factors are defined as independent. Finally, users of CFA with models that include orthogonal factors, including one weak factor, must be cautious in specifying the number of factors in the model and should know that an erroneous specification in which one factor is omitted will result in a very poor recovery of the weak factor loadings, especially if the sample size is small and the model discrepancy is suspected to be large.

As is the case with any Monte Carlo simulation study, these results will hold only in conditions similar to those considered here. Thus, future research should continue examining these effects under different study conditions. For instance, a limitation of this study is that the data were generated from a multivariate normal distribution. Thus, further study could be directed to examining whether the results presented here would hold under violations of multivariate normality. Moreover, given that the particular manner in which the misspecified models were formulated in Study 2 probably had an impact on the results, other studies should continue examining the present effects by defining structural error in other ways. Finally, another potential line of research could examine the recovery of weak factor loadings in the context of structural equation modeling, which involves a more complex situation than either EFA or CFA.

AUTHOR NOTE

This work was partially supported by Grants CCG08-UAM/ESP-3951 and CCG07-UAM/ESP-1615 from the Universidad Autónoma de Madrid and the Comunidad de Madrid (Spain). I thank Dona L. Coffman for providing the software to simulate the misspecified covariance matrices, and Javier Revuelta and Albert Maydeu-Olivares for their insightful comments. I also thank three anonymous reviewers for their helpful comments that have contributed to the improvement of the manuscript. Correspondence concerning this article should be directed to C. Ximénez, Universidad Autónoma de Madrid, Departamento de Psicología Social y Metodología, Cantoblanco s/n. 28049 Madrid, Spain (e-mail: carmen.ximenez@uam.es).

REFERENCES

- ANDERSON, J. C., & GERBING, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, **49**, 155-173.
- BOOMSMA, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Vol. I, pp. 148-173). Amsterdam: North-Holland.
- BRIGGS, N. E., & MACCALLUM, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, **38**, 25-56.
- BROWNE, M. W., & CUDECK, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- CUDECK, R., & BROWNE, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika*, **57**, 357-369.
- FAN, X., & SIVO, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, **12**, 343-367.
- FAVA, J. L., & VELICER, W. F. (1996). The effects of underextraction in factor and component analyses. *Educational & Psychological Measurement*, **56**, 907-929.
- HAKSTIAN, A. R., ROGERS, W. T., & CATTELL, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research*, **17**, 193-219.
- 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. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32**, 443-482.
- JÖRESKOG, K. G., & SÖRBOM, D. (1981). *LISREL: Analysis of linear structural relationships by the method of maximum likelihood* (Ver. V). Chicago: National Educational Resources.
- JÖRESKOG, K. G., & SÖRBOM, D. (1996a). *LISREL 8: User's reference guide* (2nd ed.). Lincolnwood, IL: Scientific Software International.
- JÖRESKOG, K. G., & SÖRBOM, D. (1996b). *PRELIS 2: User's reference guide* (3rd ed.). Chicago: Scientific Software International.
- LA DU, T. J., & TANAKA, J. S. (1989). Influence of sample size, estimation method, and model specification on goodness-of-fit assessments in structural equation models. *Journal of Applied Psychology*, **74**, 625-635.
- LEVINE, M. S. (1977). *Canonical analysis and factor comparison*. Beverly Hills, CA: Sage.
- MACCALLUM, R. C. (2003). 2001 presidential address: Working with imperfect models. *Multivariate Behavioral Research*, **38**, 113-139.
- MACCALLUM, R. C., BROWNE, M. W., & CAI, L. (2007). Factor analysis models as approximations. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 153-175). Mahwah, NJ: Erlbaum.
- MACCALLUM, R. C., WIDAMAN, K. F., PREACHER, K. J., & HONG, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, **36**, 611-637.
- McFADDEN, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105-142). New York: Academic Press.
- OLSSON, U. H., TROYE, S. V., & HOWELL, R. D. (1999). Theoretical fit and empirical fit: The performance of maximum likelihood versus generalized least squares estimation in structural equation models. *Multivariate Behavioral Research*, **34**, 31-58.
- SKRONDAL, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, **35**, 137-167.
- STEIGER, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, **25**, 173-180.
- THURSTONE, L. L. (1930). The learning function. *Journal of General Psychology*, **3**, 469-493.
- TUCKER, L. R. (1951). *A method for synthesis of factor analysis studies* (Tech. Rep. 984). Washington, DC: Department of the Army.
- TUCKER, L. R., KOOPMAN, R. F., & LINN, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, **34**, 421-459.
- VERNON, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen.
- XIMÉNEZ, C. (2006). A Monte Carlo study of recovery of weak factor loadings in confirmatory factor analysis. *Structural Equation Modeling*, **13**, 587-614.

(Manuscript received November 7, 2008;
revision accepted for publication May 5, 2009.)