In this article, the authors examine the most common type of improper solutions: zero or negative error variances. They address the causes of, consequences of, and strategies to handle these issues. Several hypotheses are evaluated using Monte Carlo simulation models, including two structural equation models with several misspecifications of each model. Results suggested several unique findings. First, increasing numbers of omitted paths in the measurement model were associated with decreasing numbers of improper solutions. Second, bias in the parameter estimates was higher in samples with improper solutions than in samples including only proper solutions. Third, investigation of the consequences of using constrained estimates in the presence of improper solutions indicated that inequality constraints helped some samples achieve convergence. Finally, the use of confidence intervals as well as four other proposed tests yielded similar results when testing whether the error variance was greater than or equal to zero.

Improper Solutions in Structural Equation Models

Causes, Consequences, and Strategies

FEINIAN CHEN
KENNETH A. BOLLEN
University of North Carolina at Chapel Hill
PAMELA PAXTON
Ohio State University
PATRICK J. CURRAN
University of North Carolina at Chapel Hill
JAMES B. KIRBY
Agency for Healthcare Research and Quality

The frequency with which improper solutions appear to occur is rather surprising.

—Joreskog and Lawley (1968:90)

AUTHORS' NOTE: *This article was previously presented at the 1999 annual meeting of the Psychometric Society at the University of Kansas, Lawrence, Kansas, June 24-27, 1999.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 29 No. 4, May 2001 468-508

We refer to improper solutions as estimates that take on values that would be impossible for the corresponding parameters or that are constrained to the boundaries of possible values. These primarily take the form of a correlation greater than one or constrained at one or a variance that is negative or constrained to zero. The type of improper solutions that have received the greatest attention are negative or constrained to zero-error variances, sometimes called *Heywood cases*. These concern us to the degree that they indicate something wrong with our model, data, or estimator. As Joreskog and Lawley's (1968) quote suggests, negative error variances occur more frequently than we might expect. Yet, despite their frequency, our knowledge about them is modest. The growing use of confirmatory factor analysis and structural equation models (SEMs) makes understanding improper solutions a pressing issue.

There are three general questions to ask about improper solutions: (1) What makes them more likely to occur? (2) What are the consequences of having them? (3) What strategies should we follow to cope with them? In this article, we address each of these questions. We highlight those areas in which our study corroborates or runs counter to prior research, and we provide new results and approaches to improper solutions. Specifically, we examine the contribution of sample size and model misspecification to the incidence of improper solutions. Key questions are whether improper solutions are a symptom of a misspecified model (omitted paths) and whether the number of improper solutions will tend to increase as the extent of model misspecification increases. Although misspecification is regarded as one of the major causes of improper solutions (Bollen 1989, Dillon, Kumar, and Mulani 1987, Van Driel 1978), we have located no studies that have examined what happens to the number of improper solutions as the degree of misspecification in a model increases. The main exceptions are studies of factor analysis models that "overfactor" their data and look at the occurrence of improper solutions (e.g., Sato 1987). A unique contribution of our article is the inclusion of both correct and incorrect model specifications examined across two sample sizes. We focus on a particular type of model misspecification in this article—omission of paths—and investigate how the number of omitted paths is related to the occurrence of improper solutions. Furthermore, we examine the consequences of improper solutions for the parameter estimates, estimated asymptotic standard errors, and chi-square statistic under three conditions: the improper solutions are freely estimated, constrained to zero, or have an inequality constraint imposed. Each of these strategies appears in research, and our goal is to provide evidence on which is the most reasonable approach to use in applied research settings.

Finally, we explore significance tests to determine whether an improper solution departs from a proper solution in a statistically significant way. This is salient because it has the potential of demonstrating that the improper solution is due to sampling fluctuations rather than model misspecification. We will follow Van Driel's (1978) suggestion of using the asymptotic standard errors of the error variances to form confidence intervals and compare it to several alternative tests of statistical significance that we suggest. Our goal is to examine the behavior of Van Driel's and our significance tests for practical use with improper solutions. It would be ideal if these tests work since we could then determine whether the cause of Heywood cases is due to sampling fluctuation, a situation of less concern than is misspecification of the model.

PAST RESEARCH

CAUSES OF IMPROPER SOLUTIONS

Van Driel (1978) wrote about improper solutions in exploratory factor analysis, but his advice is useful for SEMs in general. Specifically, he argued that improper error variance estimates might result from any one of three causes: (1) sampling fluctuations, (2) model misspecification to the extent that no factor analysis model will fit the data, and (3) "indefiniteness" (underidentification) of the model. To this list we add (4) empirical underidentification (Rindskopf 1984) and (5) outliers/influential cases (Bollen 1987). Furthermore, we expand cause (2) by considering how misspecification can shift the error variances and their standard errors, which in turn affects the probability of observing improper solutions. In other words, model misspecification may not be severe enough to create zero or negative error variances in the population, but the slight shift in the error

variance in combination with changes in sampling error could influence the chances of getting nonpositive error variances in a sample. Strictly speaking, the "direct" cause of improper solutions could still be sampling fluctuation, but we consider this type of improper solutions as a consequence of misspecification because the probability of observing improper solutions is changed. In addition, the shape of the distribution of the estimator of error variances (e.g. normal or nonnormal) can also affect the probability of nonpositive error variance estimates. Thus, our definition of *cause* of improper solutions is broader than some researchers (see Luijben 1989). Although all five of these influences are of interest, we will primarily focus on two: sampling fluctuations and model misspecification.

While the role of misspecification in generating improper solutions has been raised in earlier theoretical discussions (Bollen 1989, Dillon et al. 1987, Van Driel 1978), we could find no Monte Carlo simulation studies that investigated whether improper solutions were more frequent in misspecified models. The main exception was exploratory factor analysis research that looks at the consequences of "overfactoring" for improper solutions (Rindskopf 1984). These studies of the inclusion of too many factors might be the source of the general tendency for researchers to suspect model specification errors when the solution is improper and to treat "proper estimates" as a partial validation of a model. Our study will empirically examine the reasonableness of this approach.

Boomsma (1983, 1985) used a Monte Carlo simulation design to explore the circumstances and frequency in which nonconvergence and improper solutions may occur. She found that improper solutions occurred more frequently in small sample sizes, with smaller population values of error variances both across comparable models and within a single model, and in six-variable factor analysis models compared with eight-variable ones. She recommended that analysts avoid sample sizes of less than 50. Boomsma's study was the only one that we have found that included two full SEMs rather than just the factor analysis model that all others have relied on.

Anderson and Gerbing (1984) assessed the effect of sampling errors and model characteristics on the occurrence of nonconvergence, improper solutions, and distribution of goodness-of-fit indices in maximum likelihood (ML) confirmatory factor analysis. Consistent with findings by Boomsma (1985), they found that sample size and the number of indicators per factor were associated with the occurrence of improper solutions. They recommended a sample size of at least 150 and three or more indicators per factor in a confirmatory factor analysis.

In sum, empirical evidence suggests that sample size and model specification aspects such as the number of indicators and having too many factors are related to the incidence of improper solutions. However, previous studies have not examined typical misspecifications such as the omission of paths and their contribution to the incidence of improper solutions. In addition, general SEMs have been less studied than factor analysis models. In this article, we will address these issues.

CONSEQUENCES AND STRATEGIES

In a second study, Gerbing and Anderson (1987) studied the consequences of improper solutions in sample sizes of 75 and 150, since at larger sample sizes improper solutions were relatively rare. In their three-factor models, each factor was measured with two or three indicators, and they varied the magnitude of the loadings. They found that the loadings in the locality of the improper solutions were biased. More specifically, those factor loadings for an indicator with an improper solution tended to be positively biased, other loadings for indicators of the same factor on average underestimated the population parameter, and the remaining loadings were practically unbiased. Gerbing and Anderson also noted a general tendency for the standard errors to be overestimated for all the factor loadings, with there being greater differences for the loadings associated with the indicator with the negative error variance. However, chi-square likelihood ratio tests and some other goodness-of-fit indices were largely unaffected.

Gerbing and Anderson (1987) also investigated the consequences of constraining an error variance to be nonnegative in the estimation, fixing a negative error variance to zero, and fixing the negative error variance to a value that would lead the indicator to have about 28 percent residual variance. They discovered that the overall fit measures were not greatly affected by the unconstrained versus the constrained versions of the model, with the possible exception of the last constraint

in which the error variance was set to a positive number.[1] The loadings for the indicator with the improper solutions decreased, those for the other indicators loading on the same factor increased, and the remaining loadings were largely unaffected across the unconstrained and constrained solutions.

One of their findings is particularly relevant to Van Driel's (1978) suggestion that analysts use the estimated asymptotic standard error for the negative error variance to form a confidence interval to check to see if it includes zero. If it includes zero, Van Driel suggested that the researcher can conclude that the improper estimate is a result of sampling fluctuations rather than specification errors. The appropriateness of this test depends on the estimated asymptotic standard error being a reasonable estimate of the standard deviation of the corresponding improper estimates so that a confidence interval using it would be accurate. Gerbing and Anderson (1987) provided the only Monte Carlo simulation evidence on this issue that we could locate. They found that for the correctly specified model, the estimated asymptotic standard errors for the unique variance estimates led to no instance out of 100 samples in which the 95 percent confidence intervals did not include zero. Since in the population, the error variance parameter was greater than zero, this check did not give any false indication that the error variance was negative. However, we cannot say much more about the accuracy of the confidence interval, nor does this provide much guidance on testing for negative error variances when there are more than one error variance to test.

SIGNIFICANCE TESTS FOR ZERO AND NEGATIVE ERROR VARIANCES

Several aspects of the consequences and treatment of improper solutions require further investigation. Of key interest is the testing of whether an improper estimate is due to sampling fluctuations. We propose several other means besides the Van Driel (1978) confidence intervals to test whether the improper estimate is a result of specification error or sampling error. First, we can form the ratio of the parameter estimate to its estimated asymptotic standard error to form a test statistic to compare to a standardized normal variable. The main difference between the z test and the confidence interval method is that it is a one-tailed instead of a two-tailed test. A closely related significance test is a Wald test for the negative error variance in which the ratio of the parameter estimate to its asymptotic standard error is squared and the resulting test statistic compared to a chi-square distribution with one degree of freedom. While the Wald test coincides with the z test for an individual parameter, it can be extended to test multiple parameters, whereas the z test is not a simultaneous test. A possible problem with both of these tests is that they rely on the estimated asymptotic standard errors, and it may not be a good estimate of the standard deviation of the estimate. We will directly examine this issue here.

Another test statistic is a likelihood ratio test statistic. We can form it by comparing the likelihood value for the model with the unconstrained error variance to one that sets it to zero (or some other small positive value). The chi-square distribution is the reference distribution for this test statistic. Finally, a Lagrangian multiplier ("modification index") test statistic is another possibility. This would require only the estimation of the constrained to zero-error variance model. The Lagrangian multiplier statistic for releasing the constraint leads to a chi-square test statistic. An advantage of the Wald, Lagrangian multiplier, and likelihood ratio tests is that they all generalize to simultaneous tests of several error variances. This is preferable to repeated checks of the different error variances using multiple confidence intervals or multiple z tests.

In sum, we describe five tests for whether a negative error variance is due to sampling fluctuations: (1) a test to determine whether the confidence intervals include zero, (2) a z test of the null hypothesis that the error variance is zero versus the alternative that it is less than zero, (3) a Wald test, (4) a likelihood ratio chi-square test, and (5) a Lagrangian multiplier test. Van Driel (1978) proposed the first test, while we suggested the others here to compare to the Van Driel confidence interval method. A comparison of all these tests is important because they will reveal whether we have a practical test of whether sampling fluctuations are the cause of an improper solution.

Our goal for this article is to address three questions relating to improper solutions in SEM: (1) What makes improper solutions more likely to occur? (2) What are the consequences of having them? (3) What strategies should we follow to cope with them? To address the first question, we will use a computer simulation design to examine the individual and joint influences of sample size and model specification on the incidence of improper solutions. To address the second question, we will examine bias in resulting parameter estimates, standard errors, and test statistics. Finally, to address the third question, we will examine a variety of statistical tests to help determine whether an improper solution might be attributable to sampling fluctuation or if instead it may denote more serious model specification problems.

DESIGN

The most challenging aspect of any computer simulation study is in the careful selection of the population models. We chose specifications that reflected prototypical model types in the social science literature combined with considerations of statistical power to reject a model with a given misspecification at a given sample size. Selected parameter values led to a range of effect sizes (e.g. communalities and R^2 values ranging from 49 to 72 percent) and, for the misspecified conditions, both a wide range of power to detect the misspecifications (e.g., power ranging from .07 to 1.0 across all sample sizes) and a range of bias in parameter estimates (e.g., absolute bias ranging from 0 to 37 percent). See Paxton et al. (forthcoming) for a comprehensive description of our model parameterization. We believe that this parameterization reflects values commonly encountered in applied research and that the omission of one or more parameters would result in meaningful impacts on parameter estimation and overall model fit.

We chose misspecifications according to the degree that they would elevate the magnitude of the ML fitting function (Joreskog and Sorbom 1993) evaluated at the population covariance matrix of the observed variables. The ML fitting function is zero in the population for a correct model and increases in magnitude as we misspecify the model by removing paths.[2] In this article, by degree of misspecification, we specifically refer to the number of paths that have been omitted from the correctly specified model. We do not consider other types of misspecification, such as inclusion of nonexistent paths or misspecification of error structure.

Our models are presented in Figures 1 and 2. Model 1 is a causal chain model with three latent variables (η_1, η_2, and η_3). Each latent factor has three indicators. Among the nine observed variables (Y1-Y9), three variables (Y4, Y6, and Y7) have cross loadings: Y4 is an effect indicator of both η_1 and η_2, Y6 is an effect indicator of both η_2 and η_3, and Y7 is an effect indicator of both η_2 and η_3. The three scaling indicators, λ_{21}, λ_{52}, and λ_{83} (underlined in Figure 1) are fixed at 1.0 in unstandardized metric. The population values of the lambdas are all set to be 1.0 in unstandardized metric, except for the cross loadings, which are set to be 0.3. Population values of β_{21} and β_{32} are set to be 0.6. The numbers in parentheses below the unstandardized coefficients are the standardized coefficients. The error variances are set to values so as to achieve R^2s of 0.49.

The measurement component of model 2 is the same as in model 1. This model also includes four exogenous variables, measured without error. All four of the exogenous variables affect the first latent variable, η_1. The first and third exogenous variables, X1 and X3, also affect η_2 and η_3. We selected the population parameter values of the covariances between the exogenous variables and the paths from the exogenous variables to the latent variables to maintain the parameter values of the measurement model.

Past research has not focused on the interplay between misspecification and negative error variances. Therefore, for each model we estimated four model specifications—three of which were misspecifications. That is, in three of the specifications, the model estimated in the sample did not correspond to the model in the population. For model 1, specification 1 is the correctly specified model, corresponding to the model picture in Figure 1. In specification 2, we omitted the path from η_2 to Y7 (henceforth λ_{72}). λ_{72}, along with the other parameters that will be omitted in specifications 2 through 4 are dashed in Figure 1 for easier identification. Specification 3 additionally omits the path from η_3 to Y6 (λ_{63}), while specification 4 additionally omits the effect from η_1 to Y4 (λ_{41}). In model 1, consequently, the degree of misspecification as gauged by the fitting function value increases from specification 2 to 4, with first one, then two, then three paths removed from subsequent specifications.

For model 2, specification 1 is the correctly specified model, corresponding to Figure 2. In specification 2, we omitted all three of the

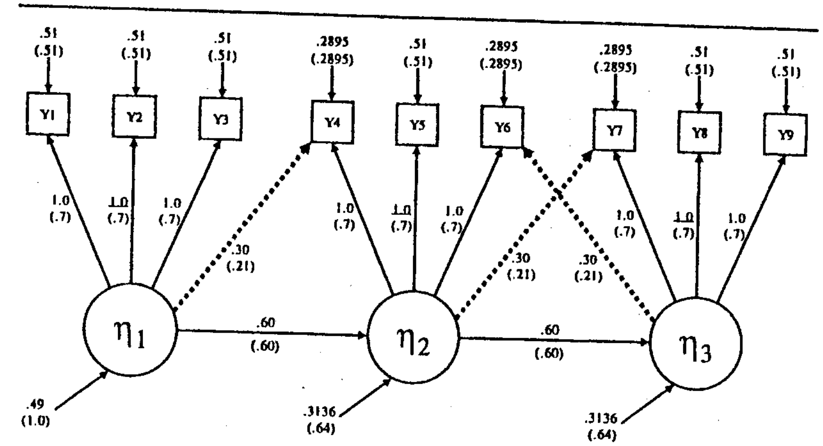

Figure 1: Diagram of Model 1

paths removed in model 1: λ_{41}, λ_{63}, and λ_{72}. In specification 3, only the four gamma paths are omitted, X1 to η_2 (γ_{21}), X3 to η_2 (γ_{3}), X1 to η_3 (γ_{31}), and X3 to η_3 (γ_{33}). In specification 4, both the four gamma and the three lambda paths are omitted. Like the previous model, we chose the ordering of these misspecifications so as to increase the magnitude of the fitting function as we remove paths. Finally, none of the misspecified models results in negative population error variances. Thus, any sample negative error variances are due to other influences.

EQS (Bentler 1997) was used for data simulation. Our models define the population covariance matrix, which we used to generate sets of raw data. From a multivariate normal distribution, we obtained 650 sets of raw data from the population with $N = 50$ and $N = 75$. We used 100 as the maximum number of iterations. Nonconvergence was rare (e.g., ranging from 0 to 5 samples across different specifications for model 1, $N = 50$, using EQS). For the analysis in this article, we only used converged cases. See the technical appendix (http://www.unc.edu/~curran/csim.html) for the exact number of samples used for each model. Our focus is on small sample sizes ($N =$ 50, 75) since improper solutions were rare for our sample size at or above 100, and the greatest potential for bias appears to be in small sample sizes. For example, for model 1, specification 1 at $N = 100$, we

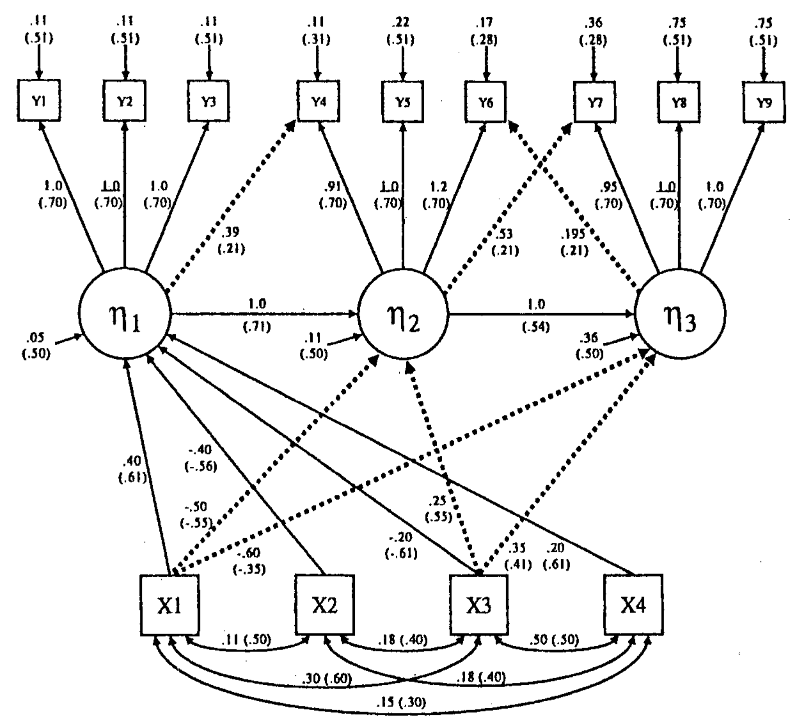

Figure 2: Diagram of Model 2

found only 12 out of 650 samples with nonpositive error variances. At larger sample sizes, they were even less frequent.

We estimated each of the eight (two models × four specifications each) specifications on the samples of raw data, obtaining parameter estimates and fit statistics. To compare estimation with negative error variance restrictions to unrestricted estimation, we estimated all models in both EQS and SAS's PROC CALIS. We used EQS for constrained estimation and CALIS for unconstrained estimation.

RESULTS

In our investigation of improper solutions, we consider the number of improper solutions, the bias in the parameters introduced by the

presence of improper solutions, and the difference in bias from constrained versus unconstrained estimation. We then discuss briefly the comparison between asymptotic standard errors and empirical standard deviation in these different situations. We also investigate the behavior of Van Driel's (1978) significance test and our four additional tests of the statistical significance of the negative error variances. Within our design, there are four conditions that vary: model (model 1 or model 2), misspecification (specifications 1-4), sample size (50 or 75), and estimation technique (constrained or unconstrained). With this large number of conditions (two models, four specifications, two sample sizes, and two estimation techniques), we produced voluminous results. To conserve space and retain focus in the present article, we present the figures and tables that help illustrate our findings. Interested readers can refer to the technical appendix (http://www.unc.edu/~curran/csim.html) for further details. Examples of the tables and figures in the technical appendix include full comparisons of parameter estimates among all samples, just samples with proper solutions, and just samples with improper solutions across all conditions; side-by-side comparisons of the constrained and unconstrained estimation techniques; and box plots to visually display differences in the constrained and unconstrained estimation techniques.

INCIDENCE OF IMPROPER SOLUTIONS

Interestingly, our results indicate that in model 1, the number of zero or negative error variances *decreases* as the number of omitted paths *increases* (see Figure 3). This finding holds across both sample sizes. In model 1, specification 1 (i.e., properly specified), $N = 50$, 15.2 percent (99) of the 650 replications were samples with improper solutions.[3] For specification 2 (λ_{72} path omitted), only 11.1 percent (72) of the replications had improper solutions. The number of improper solutions decreased slightly, to 10.0 percent (65), for specification 3 (λ_{72} and λ_{63} omitted) and then to 8.3 percent (54) for specification 4 (λ_{72}, λ_{63}, and λ_{41} omitted).

Continuing to examine model 1, at $N = 75$, we see fewer improper solutions than at $N = 50$. This is consistent with past studies that suggest an inverse relationship between sample size and frequency of

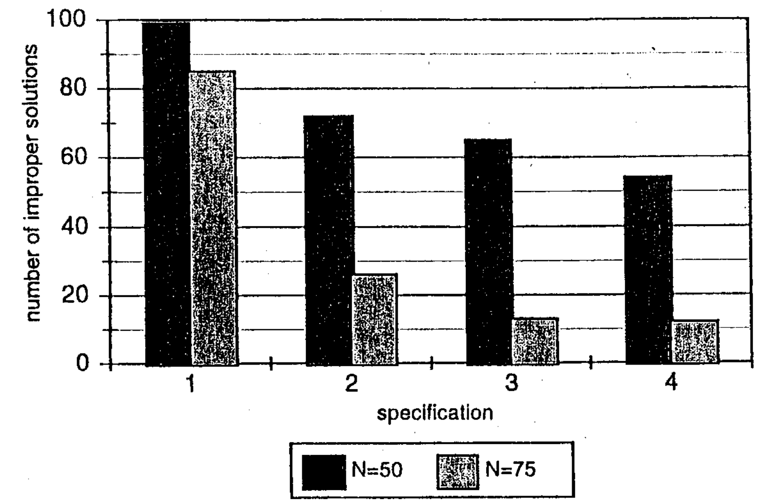

Figure 3: Number of Improper Solutions: Model 1

improper solutions (Boomsma 1985). With the larger sample size, we again see that the degree of misspecification is inversely related to number of improper solutions. For specification 1, 13.1 percent (85) of the samples had improper solutions, which reduced to 4.0 percent (26) at specification 2, 2.0 percent (13) for specification 3, and 1.8 percent (12) for specification 4.

For model 2 (displayed in Figure 4), the relationship between the number of negative error variances and the degree of misspecification is less straightforward than in model 1. Recall that we are measuring the degree of misspecification by the magnitude of the ML fitting function for a given model evaluated at the population covariance matrix. Our specifications for model 2, like model 1, are ordered from a zero fitting function value for specification 1 to monotonically increasing values as we move to specifications 2, 3, and 4. The power of the chi-square tests for these models corresponds to the same ordering. This was discussed in the Design section.

In model 2, specification 1, $N = 50$, 83 (12.8 percent) of the samples had improper solutions. This number sharply decreased to 31 (4.8 percent) when all three λ paths (λ_{72}, λ_{63}, and λ_{41}) were omitted in specification 2. At this point, the findings mirror those of model 1—when the

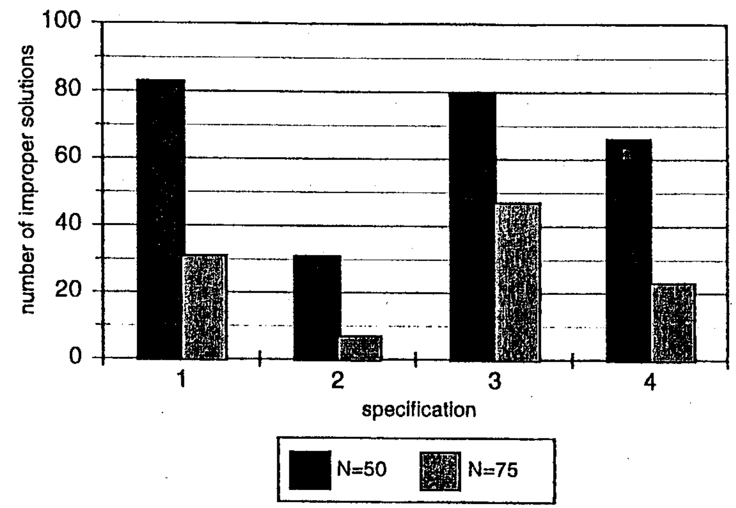

Figure 4: Number of Improper Solutions: Model 2

measurement model is misspecified, the number of improper solutions decreases. However, in specification 3, when only the gamma paths (γ_{21}, γ_{23}, γ_{31}, and γ_{33}) are omitted, we find no decrease in the number of improper solutions and in fact find an increase relative to specification 2. With both the lambda and gamma paths omitted in specification 4, the number of improper solutions was 66 (10.2 percent)—fewer than the properly specified model but more than specification 2.

With $N = 75$, we observe similar results. Again, there are fewer improper solutions at the larger sample size. The number of improper specifications at specifications 1 through 4 are 31 (4.8 percent), 7 (1.1 percent), 47 (7.2 percent), and 23 (3.5 percent) respectively. With the lambda paths removed (specification 2), there are the fewest improper solutions. At this sample size, however, when the gamma paths are removed there are more improper solutions than the correctly specified model.

In sum, in contrast to model 1, in model 2 there is no monotonic decline in the number of improper solutions with increasing misspecification. Instead, the omission of lambda paths seems to decrease improper solutions, while the omission of gamma paths increases the number of improper solutions. Omission of both lambda and gamma paths falls somewhere in between. In both models, however, misspecifications that omit lambda paths in the measurement model appear to decrease the number of improper solutions.

It seems that improper solutions are more frequent for error variance parameters that are closer to zero than for those that are further away from zero. Evidence relevant to this question comes from Figure 5. It contains the number of negative error variances by parameter for model 1, specification 1 for $N = 50$. Consistent with the idea that improper solutions will be more frequent for error variances that are closer to zero in the population is that Y7, Y6, and Y4 have the lowest error variances (all equal 0.29), and these also are the error variances with the greatest frequencies of negative values. However, the lower error variances cannot be the entire explanation, since Figure 5 shows that the number of improper solutions is not the same for these three variables despite the fact that they have equal population error variances. Furthermore, these results are not unique to model 1, specification 1. The same pattern holds for the other specifications for model 1 and model 2.

Our preceding findings reveal that the relation between the magnitude of misspecification (i.e., the value of F_{ml} in the population) and the frequency of improper solutions is not a simple association. Misspecification can make nonpositive error variances more likely in several ways. One is if the misspecification leads some error variances to be nonpositive in the population. This is one sense in which we would say that misspecification caused an improper solution in the population and this negative value typically increases the probability of improper solutions in the sample estimates. However, misspecifications can affect the probability of improper solutions in samples even if the error variance is positive in the population. To understand how, we need to consider that the probability of a nonpositive error variance is affected by several factors. One part is the population parameter value of the error variance and how close it is to zero. If positive, the closer it is to zero, the greater the probability of a nonpositive error variance, other things equal. A second factor is the standard deviation of the estimator of the error variance. Greater standard deviations suggest a wider dispersion of estimates and hence a greater probability of nonpositive error variances than when smaller standard deviations occur. Finally, the shape of the distribution of the

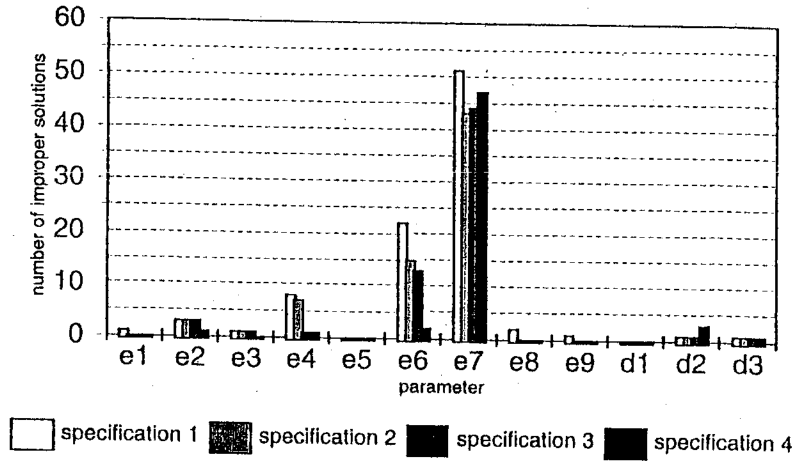

Figure 5: Improper Solutions by Parameter: Model 1, $n = 50$

estimator of the error variances will affect the probability of nonpositive error variance estimates. For instance, even if the population error variance and the standard deviation of the estimator are identical in two situations, a left-skewed distribution for an estimator of variance could have a higher probability of nonpositive error variance than a normal distribution.

We partially investigated these factors influencing the probability of negative error variances for our simulation models. First, we fit correct and incorrect specifications to the population covariance matrices for models 1 and 2. In no case did the misspecification lead to population parameters that were nonpositive. Hence, all the negative error variance estimates are due to sampling fluctuations.[4] To further investigate this, we looked at the population error variance for E6 for model 1 under all four specifications, and these were 0.29, 0.29, 0.24, and 0.29. The population error variance for the misspecified models differs little from the correct specification. If anything, this difference alone would not lead us to predict a greater probability of nonpositive error variances in the misspecified models (with the possible exception of specification 3) than in the correct specification. The fact that the number actually decreases in some misspecified models suggests that we need to consider the mean, standard deviations, and shape of the distributions for these error variances in the simulation samples.

Figure 6 provides the histograms for the error variance estimates for E6 across the four specifications, based on our simulated samples. The mean error variances are 0.24, 0.25, 0.23, and 0.28 from specification 1 to specification 4. These are consistently lower than the population error variances reported in the previous paragraph. The empirical standard deviations are 0.16, 0.14, 0.11, and 0.11, respectively. There are also more extreme cases in specification 1 and 2 compared with the other specifications. A slightly smaller mean error variance and a bigger standard deviation for the correctly specified model partly explain why we get more nonpositive variances in this model compared with other misspecified models. In addition, the shape of the distributions appears more "normal" in more misspecified models than the correctly specified model. For example, the histogram for specification 1 is left skewed and thin looking (skewness = –5.84, kurtosis = 69.49). The histogram for specification 4 is almost normal looking (skewness = .30, kurtosis = .452). In sum, changes in mean error variance, standard deviation, and shape of the distribution led to a drop in improper solutions for E6 from specification 1 (22 cases) to specification 4 (2 cases) (see Table 1e in technical appendix).

In sum, the results of this section indicate the hazard of considering improper solution estimates as a straightforward indicator of model misspecification. As our results illustrate, misspecifications of a model can decrease the probability of improper solution estimates. Of course, if we had the population covariance matrix and found nonpositive error variances, then we would have clear evidence of misspecification. However, in the real-world situation of sample data, a negative error variance estimate can occur due to sampling fluctuations around a positive error variance in the population. Thus, it is important to determine whether a nonpositive error variance is statistically significant. We return to this issue below.

CHI-SQUARE TEST STATISTIC

We now examine the consequences of improper solutions. Table 1 presents the mean chi-square test statistic for all four specifications of models 1 and 2 for the proper and improper solutions and for the

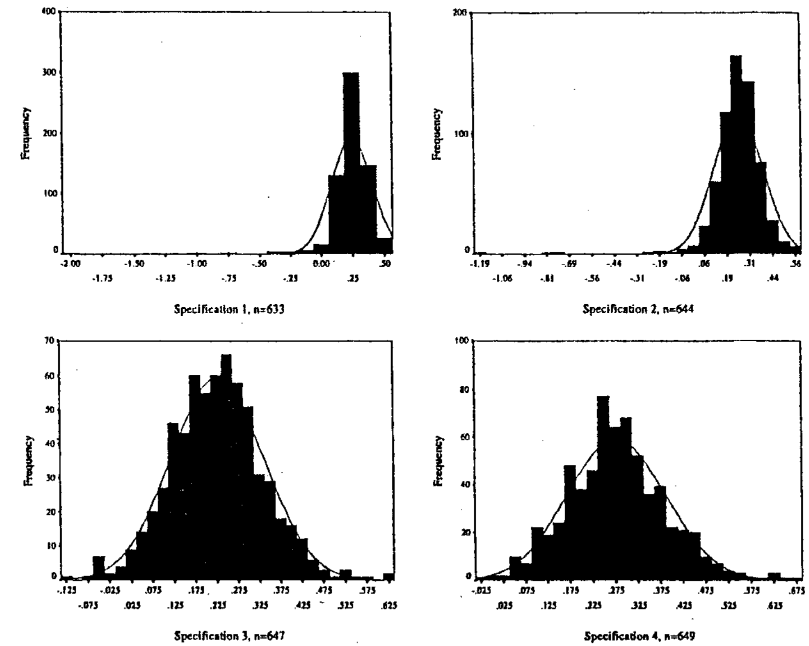

Figure 6: Histograms of E6, Model 1, n = 50, Specification 1-4

constrained and unconstrained samples. To examine the impact of improper solutions on the test statistic, we can compare its mean value for the proper and improper solutions for the same model and same specification. Doing so, we find little practical difference in the test statistic, replicating the findings of Anderson and Gerbing (1984).

BIAS IN PARAMETER ESTIMATES

If we compare the parameter estimates for samples that include only proper solutions with samples that include only improper solutions, the samples with improper solutions reflect the largest bias. Table 2 compares mean relative bias in the parameter estimates for samples with improper solutions against those with only proper solutions for both models at sample size 50 under constrained estimation.[5]

TABLE 1: Comparisons of Chi-Square Statistics Between Samples With Only Proper and Only Improper Solutions

	Model 1			Model 2		
		Improper			Improper	
	Proper	Constrained	Unconstrained	Proper	Constrained	Unconstrained
$N = 50$						
Specification 1	($df = 22$)			($df = 50$)		
Null chi-square	212.33	214.50	217.09	361.74	357.16	359.37
Model chi-square	23.72	24.76	24.56	57.67	57.96	57.41
N	546	99	86	560	83	76
Specification 2	($df = 23$)			($df = 53$)		
Null chi-square	212.52	213.92	214.57	360.33	369.31	369.31
Model chi-square	25.89	27.04	25.77	67.24	67.11	66.72
N	576	72	68	618	31	31
Specification 3	($df = 24$)			($df = 54$)		
Null chi-square	212.82	212.31	212.12	361.86	354.17	356.21
Model chi-square	27.77	29.19	28.33	81.50	79.51	78.93
N	584	65	63	567	80	70
Specification 4	($df = 25$)			($df = 57$)		
Null chi-square	212.93	209.94	211.09	360.80	360.67	360.67
Model chi-square	31.69	31.37	31.01	91.35	94.11	93.74
N	596	54	54	583	66	66
$N = 75$						
Specification 1	($df = 22$)			($df = 50$)		
Null chi-square	302.70	289.56	289.95	491.81	478.50	473.94
Model chi-square	23.15	24.16	24.41	54.67	53.19	52.59
N	610	38	37	618	31	28
Specification 2	($df = 23$)			($df = 53$)		
Null chi-square	302.58	287.06	289.75	491.46	467.47	467.47
Model chi-square	25.53	26.20	26.40	66.87	64.39	64.11
N	624	26	24	643	7	7
Specification 3	($df = 24$)			($df = 54$)		
Null chi-square	302.08	295.66	295.66	492.16	478.99	480.83
Model chi-square	27.84	32.54	32.12	87.99	87.23	85.89
N	637	13	13	603	47	45
Specification 4	($df = 25$)			($df = 57$)		
Null chi-square	302.12	293.24	293.24	491.22	490.89	492.34
Model chi-square	32.78	38.22	37.80	101.84	101.08	100.48
N	638	12	12	627	23	24

Beginning with model 1, $N = 50$ (the left top half of Table 2), it is useful to trace the pattern of bias introduced by misspecification in the properly specified model and proper solutions. We define relative bias as 100 multiplied by the difference between the parameter estimate and its population value divided by the population value. As the lambda paths are removed, bias rises in the related cross-loading

estimates. For example, when λ_{72} is set to zero in specification 2, the mean relative bias in λ_{73} rises from 1.8 percent to 33.2 percent. In specification 4, when λ_{41}, λ_{63}, and λ_{72} are removed, λ_{42}, λ_{62}, and λ_{73} show biases of 23.5 percent, 26.3 percent, and 33.7 percent, respectively. As for the structural paths, the estimates of β_{21} show little change over the misspecifications, while β_{32} shows generally increasing bias with misspecification –0.6 percent to 11.5 percent to 28.4 percent to 24.9 percent. In general, in the samples with proper solutions, we see increasing mean relative bias with increases in misspecification. We also see more parameter estimates experiencing increases in bias when the number of specification errors is increased. These findings are consistent with our theoretical expectations.

Moving to the samples with improper solutions, a direct comparison of the proper/improper samples shows that the samples with improper solutions tend to have larger bias in the parameter estimates.[6] In fact, the bias increases dramatically for some parameters. Like the samples with only proper solutions, bias is most severe for the cross loadings. Even in the correctly specified model (specification 1), estimates of λ_{72} and λ_{63} show mean relative biases of –149.0 percent and –84.6 percent when negative error variances are included. The bias is also large in their surrounding factor loadings, λ_{62} and λ_{73} (30.3 percent and 65.2 percent). Biases in other parameter estimates are also relatively large. For example, λ_{41} shows a bias of –20.0 percent (compared with –.6 percent in the proper-solution sample), and β_{32} shows a bias of 22.1 percent (compared with –.6 percent).

Similar findings appear across the misspecifications—the coefficient estimates from samples with negative error variances tend to be more biased than samples with only proper solutions. This follows the pattern determined in the samples with proper solutions only. Also similar to the samples with proper solutions, as lambda paths are omitted in the misspecifications, the bias in the other cross loadings tends to increase. For example, in specification 2, the bias in the estimates of λ_{63} increases from –84.6 percent to –133.3 percent, while λ_{41}, λ_{42}, and λ_{62} estimates increase in bias from –20.0 percent, 16.4 percent, and 30.3 percent to –59.5 percent, 25.5 percent, and 51.9 percent.[7] Bias in the improper-solution samples decreases from these highs to lower levels in specifications 3 and 4 but remain higher than those in the proper-solution samples.

TABLE 2: Mean Relative Bias (percentage) in the Parameters of the Models for Samples With Only Proper or Only Improper Solutions

	Specification 1		*Specification 2*		*Specification 3*		*Specification 4*		*Specification 1*		*Specification 2*		*Specification 3*		*Specification 4*	
Parameter	*Proper*	*Improper*	*Proper*	*Improper*	*Proper*	*Improper*	*Proper*	*Improper*	*Proper*	*Improper*	*Proper*	*Improper*	*Proper*	*Improper*	*Proper*	*Improper*
	Model 1, $N = 50$, constrained estimation								Model 1, $N = 75$, constrained estimation							
V1F1PE	7.1	6.7	7.5	2.3	7.4	1.6	7.4	–0.2	3.5	–1.6	3.2	1.1	3.1	2.6	3.4	5.8
V3F1PE	6.6	3.6	6.7	0.8	6.5	–0.7	6.1	1.6	1.5	3.4	1.6	1.7	1.6	5.3	1.8	2.0
V4F1PE	–0.6	–20.0	2.1	–59.5	32.9	53.9	23.5	26.0	–5.3	–18.0	–4.5	35.1	28.9	38.7	24.1	32.7
V4F2PE	3.7	16.4	3.0	25.5	–4.7	–7.5	26.3	23.9	4.9	18.9	4.6	–6.6	–3.9	0.1	27.8	30.8
V6F2PE	4.6	30.3	0.6	51.9	30.9	35.8	—	—	5.8	50.1	2.0	50.2	33.3	33.6	—	—
V6F3PE	–6.9	–84.6	2.7	–133.3	—	—	—	—	–5.0	–105.0	5.2	–110.4	—	—	—	—
V7F2PE	–0.8	–149.0	—	—	—	—	—	—	–4.9	–264.6	—	—	—	—	—	—
V7F3PE	1.8	65.2	33.2	64.4	33.8	69.0	33.7	77.2	7.0	118.6	37.9	48.7	38.5	69.2	39.0	66.6
V9F3PE	1.7	2.2	0.7	6.3	1.0	5.2	1.0	5.0	4.2	–2.9	3.7	–2.5	3.7	–9.6	3.7	–5.4
F2F1PE	6.7	–4.7	6.2	–1.8	–1.4	–1.6	10.3	12.5	4.5	–9.0	4.5	–11.9	–3.4	–15.4	8.9	–2.3
F3F2PE	–0.6	22.1	11.5	14.1	28.1	9.0	24.9	–3.2	0.1	20.8	9.6	11.4	24.4	5.5	20.8	3.8
N	546	99	576	72	584	65	596	54	610	38	624	26	637	13	638	12
	Model 2, $N = 50$, constrained estimation								Model 2, $N = 75$, constrained estimation							
V1F1PE	1.0	–4.2	0.2	0.5	0.9	–3.0	0.2	–0.5	3.0	3.5	2.9	16.0	3.3	2.3	3.2	7.1
V3F1PE	3.0	–2.8	2.2	6.1	3.4	–5.0	2.5	5.7	2.7	6.3	2.7	3.4	3.1	2.9	3.3	–1.4
V4F1PE	–15.0	–14.6	—	—	9.5	69.5	—	—	–2.2	15.8	—	—	14.6	12.9	—	—
V4F2PE	10.1	12.1	22.1	23.5	5.8	–2.5	20.8	28.1	4.2	5.3	20.1	11.2	2.3	8.5	19.3	22.0
V6F2PE	6.7	57.9	28.4	37.1	1.7	43.4	29.2	37.8	8.3	44.0	26.8	19.8	2.7	41.7	28.1	23.9
V6F3PE	–5.8	–152.3	—	—	12.0	–110.3	—	—	–17.4	–139.3	—	—	3.1	–90.0	—	—
V7F2PE	–2.6	–266.1	—	—	–9.4	–219.5	—	—	–18.8	–175.2	—	—	–11.7	–253.0	—	—
V7F3PE	5.8	82.9	35.4	72.7	8.6	86.3	38.1	69.7	10.0	80.2	33.4	65.6	6.5	108.7	36.8	65.3
V9F3PE	3.6	0.0	3.7	–0.3	4.0	0.6	3.3	2.0	2.2	1.2	2.1	–1.9	2.6	1.6	2.2	4.0
F1V10PE	0.4	1.7	0.7	2.3	–12.8	–9.8	–13.6	–17.8	–0.7	0.4	–0.6	1.8	–13.0	–15.4	–14.7	–14.8
F1V11PE	–3.0	–2.1	–2.6	–4.3	–6.0	–5.8	–6.3	–3.4	2.6	0.4	2.3	14.5	–0.9	2.0	–1.0	–4.4
F1V12PE	–0.8	–3.3	–0.5	–12.6	12.1	10.3	14.3	14.1	–0.1	5.1	–0.2	–9.8	12.9	6.8	14.6	10.8
F1V13PE	2.6	2.5	2.2	10.3	5.7	8.3	6.0	6.4	–1.0	–10.6	–0.8	–10.7	2.3	0.5	2.3	7.5
F2V10PE	0.6	–1.6	0.2	11.8	—	—	—	—	–1.2	2.8	0.1	–14.8	—	—	—	—
F2V12PE	1.8	6.3	2.1	0.5	—	—	—	—	0.1	2.4	0.2	22.8	—	—	—	—
F3V10PE	–2.9	41.9	15.5	32.3	—	—	—	—	1.6	32.6	15.6	30.9	—	—	—	—
F3V12PE	1.8	–39.6	–16.7	–34.9	—	—	—	—	0.5	–31.3	–14.7	–38.1	—	—	—	—
F2F1PE	3.6	–5.4	4.7	10.7	–31.0	–40.1	–27.6	–19.0	5.0	–7.5	7.2	32.4	–30.0	–38.4	–25.6	–21.5
F3F2PE	2.9	40.6	29.5	14.8	11.4	31.9	38.4	27.3	4.5	23.4	25.5	0.2	10.7	27.9	34.4	9.6
N	560	83	618	31	567	80	583	86	618	31	643	7	603	47	627	23

NOTE: To see the dispersion measures of these parameter estimates (e.g., standard deviation, mean standard error, mean squared error), refer to Tables 1a-1d, 2a-2d, 3a-3d, and 4a-4d in the technical appendix). F1 is η_1, Y1F1 is λ_{11}, F2F1 is β_{21}, and F1X1 is γ_{11}, and so forth.

A few additional findings appear from an examination of model 1. First, those parameter estimates with a population value of 1.0 and factor complexity of one retain a low mean relative bias across misspecifications. That is, the estimates of λ_{11}, λ_{31}, and λ_{93} are relatively unaffected by bias throughout the specifications, even in the case of improper solutions. More surprisingly, the mean relative bias for two of those, λ_{11} and λ_{31}, have lower bias in the solutions with improper solutions than they do in the proper-only samples. A similar phenomenon occurs in β_{32}. There, bias in the parameter is higher in the improper samples for specifications 1 and 2 but falls for specifications 3 and 4—actually falling below the bias from the proper-solution samples. We observed a similar pattern with $N = 75$ (see the top right part of Table 2).

We next turn to model 2 (see bottom half of Table 2).[8] As mentioned earlier, model 2 is an expansion of model 1 such that four observed variables, X1, X2, X3, and X4, were added to the model as exogenous factors influencing three latent factors, η_1, η_2, and η_3. This allows us to examine the behavior of coefficient estimates linking exogenous variables to endogenous variables (gammas), as well as endogenous variables to each other (betas), which were not studied in previous research. For the correctly specified model with a sample size of 50, for those 560 proper solutions, the parameter estimates show low bias. The highest bias was in λ_{41} and λ_{42}, with mean relative bias of –15.0 percent and 10.1 percent. The remaining parameter estimates have mean relative bias of less than 10 percent. As expected, parameter estimates in improper solutions have much higher bias. For the lambdas, the bias is concentrated in the estimates of λ_{62}, λ_{63}, λ_{72}, and λ_{73}, similar to model 1. For example, the mean relative bias in the estimate of λ_{72} is –266.1 percent. For the gammas, the estimates of γ_{31} and γ_{33} have the highest bias, with mean relative bias among improper solutions as high as 41.9 percent and –39.6 percent. For the beta estimates, β_{32} has a mean relative bias of 40.6 percent.

In specification 2, similar to model 1, bias in λ_{42}, λ_{62}, and λ_{73} estimates increases for proper solutions (e.g., from a mean of 6.7 percent to 28.4 percent for λ_{62}). Interestingly, the betas and gammas are also affected. Mean relative bias of γ_{31} among proper solutions increases from –2.9 percent to 15.5 percent. The bias also increases for the parameter estimates of γ_{33} and β_{32}. For the improper solutions, however, there is an increase in mean and median bias in those parameter estimates, and extreme cases decrease considerably (see Figure 3a and Figure 5b in the technical appendix).

In specification 3, where the gamma paths (γ_{21}, γ_{23}, γ_{31}, and γ_{33}) are omitted, again as expected, the parameters that surround the omitted parameter estimates are affected most. For proper solutions, there is small increase in relative bias in estimates such as those for γ_{11}, γ_{12}, and γ_{13}. There is some slight increase in bias in some estimates of lambdas, such as λ_{63} and λ_{72}. Estimates of β_{21} are also biased to –31.0 percent. For the improper solutions, the parameter estimates are similar to specification 1. For some parameter estimates, there is a slight increase in bias, such as λ_{41} and λ_{73}. In specification 4, in which the gamma and lambda paths are all omitted except for the parameters that were not affected much in either specification 2 and 3 (such as λ_{11}, λ_{31}, and λ_{14}), all the other parameter estimates increased in bias for both proper and improper solutions.

In sum, while the bias in the parameters shows a slightly complicated relationship under conditions of improper solutions and misspecification, in general researchers will increase, often significantly, the bias in their parameter estimates by using samples with negative error variances. This holds across a variety of misspecifications. This analysis also serves to remind researchers of the large increases in bias that occur with misspecifications of their models, both in samples with proper solutions and samples that include improper solutions.

ASYMPTOTIC STANDARD ERRORS

The means of asymptotic standard errors are very similar to empirical standard deviations for both the proper and improper solutions. The one exception is for the parameter estimates that have relatively high biases in improper solutions, empirical standard deviations tend to be much bigger than mean standard errors, which are higher to begin with, than their counterparts in proper solutions. For example, for model 1, specification 1, $N = 50$, mean standard error and empirical standard deviation for λ_{63} is 0.29 and 0.32, respectively, in proper solutions but as high as 0.42 and 0.68 in improper solutions. Our technical appendix (http://www.unc.edu/~curran/csim.html) has even more detailed information on this comparison.

CONSTRAINED VERSUS UNCONSTRAINED ESTIMATION

Until this point in the discussion, we have only considered constrained estimation. As discussed earlier, however, the type of estimation varies across software, and there is some debate as to whether constrained estimation (such as performed in EQS, Bentler 1997) is a reasonable response to improper solutions. Table 3 provides information about the mean relative bias for constrained versus unconstrained estimation.[9] If we undertake a direct comparison of the mean relative bias under constrained and unconstrained estimation, we see that for those parameter estimates that are severely biased under constrained estimation, the mean relative bias is even larger under unconstrained estimation. For example, in model 1, specification 1, while the estimates for λ_{72} have a mean relative bias of –149.0 percent under constrained estimation, the bias is as big as –393.7 percent under unconstrained estimation. Similarly, in specification 2, estimates of λ_{41} have a bias of –59.5 percent under constrained estimation but –153.6 percent under unconstrained estimation. While occasionally (e.g., λ_{41} in specification 3) the unconstrained estimation produces a lower mean relative bias, unconstrained estimation typically produces a higher bias. For the parameter estimates with very small bias, the constrained versus unconstrained results are not very different.

In this case, however, considering only mean relative bias could be misleading because this could be inflated due to extreme cases or a very skewed distribution. Therefore, we show the distribution of the bias in selected parameters in the forms of box plots.[10] Figure 7 presents box plots with comparison of constrained and unconstrained estimation for model 1, specification 1, $N = 50$.[11] Although the box plots only provide information about one specification and sample size, the results are similar across specifications, sample sizes, and models.

Most of the median relative biases are not far from zero, except for λ_{72} and λ_{73}. In addition, λ_{63}, λ_{72}, λ_{41}, and λ_{73} displayed larger variances, while other estimates clustered closely around the median. The results reflect that the median difference between constrained and unconstrained estimation is much smaller. Clearly, median relative biases were lower than mean relative biases for almost all estimates. For several estimates (λ_{11}, λ_{62}, and λ_{63}), the differences were excessively large. For example, the mean and median relative bias for λ_{62} is 30.3 percent and 2.7 percent, respectively, from constrained results and 66.1 percent and 1.5 percent from unconstrained results. Thus, the difference between the constrained and unconstrained median estimates was smaller than that between their mean estimates. The reason that the mean relative bias under unconstrained estimation is so much larger is that it has more extreme cases. Overall, however, constrained estimates still appeared less biased than unconstrained estimates.

Additional points to raise from Table 3 are that some of the samples with improper solutions that converge under constrained estimation do not do so with unconstrained estimation. This suggests an advantage of the constraint—it helps achieve convergence. However, convergence seems to cover another serious problem: The parameter estimates are much more severely biased in those samples that converge only under constrained estimation than the ones that converged under both types of estimation. For example, λ_{63} under model 1, specification 1, $N = 50$, is underestimated by –444.4 percent on average with constrained estimation when unconstrained estimation would not have converged. This is a much bigger bias than that found when the samples converged under both constrained and unconstrained estimation (–73.9 percent). This large difference is mainly driven by some extreme cases since the median relative bias is only –52.8 percent. Of 17 cases that converged under constrained but not unconstrained estimation, at least 5 have relative bias greater than –173.8 percent for the estimates of λ_{63}.

Of course, it is important to stress the more general result—that samples with improper solutions result in substantially higher levels of bias across models, specifications, sample sizes, and estimation techniques. Overall, the findings in this section suggest that if a sample will converge under both constrained and unconstrained estimation, then the estimates from constrained estimation will be less biased than those from the unconstrained estimation. If, however, a sample converges under constrained estimation but would not have converged

(text continues on p. 497)

TABLE 3: A Comparison of Mean Relative Bias (percentage), Constrained Versus Unconstrained Estimation for Samples With Only Improper Solutions

	Specification 1		Specification 2		Specification 3		Specification 4	
	Constrained	Unconstrained	Constrained	Unconstrained	Constrained	Unconstrained	Constrained	Unconstrained
			Model 1, $N = 50$, samples with improper solutions					
V1F1PE	6.7	5.8	2.3	1.6	1.6	1.1	−0.2	1.2
V3F1PE	3.6	3.0	0.8	0.2	−0.7	−1.1	1.6	−0.4
V4F1PE	−20.0	−22.0	−59.5	−153.3	53.9	12.7	26.0	25.7
V4F2PE	16.4	20.5	25.5	55.6	−7.5	−0.2	23.9	22.8
V6F2PE	30.3	66.1	51.9	62.6	35.8	37.7	—	—
V6F3PE	−84.6	−160.6	−133.3	−155.6	—	—	—	—
V7F2PE	−149.0	−393.7	—	—	—	—	—	—
V7F3PE	65.2	156.8	64.4	84.2	69.0	91.5	77.2	105.7
V9F3PE	2.2	1.8	6.3	8.9	5.2	6.5	5.0	8.3
F2F1PE	−4.7	−5.5	−1.8	1.3	−1.6	−3.2	12.5	15.1
F3F2PE	22.1	26.8	14.1	7.8	9.0	3.9	−3.2	−12.0
N	99	86	72	68	65	63	54	54
			Model 2, $N = 50$, samples with improper solutions					
V1F1PE	−4.2	−3.7	0.5	0.5	−3.0	−2.2	−0.5	−0.7
V3F1PE	−2.8	−3.2	6.1	6.2	−5.0	−2.9	5.7	5.9
V4F1PE	−14.6	−33.3	—	—	69.5	6.3	—	—
V4F2PE	12.1	20.8	23.5	23.7	−2.5	7.3	28.1	29.5
V6F2PE	57.9	310.0	37.1	38.2	43.4	44.5	37.8	39.5
V6F3PE	−152.3	−1018.2	—	—	−110.3	−92.8	—	—
V7F2PE	−266.1	−2566.2	—	—	−219.5	−660.8	—	—
V7F3PE	82.9	863.1	72.7	84.1	86.3	284.1	69.7	82.9
V9F3PE	0.0	1.3	−0.3	0.0	0.6	1.2	2.0	2.1
F1V10PE	1.7	0.3	2.3	2.2	−9.8	−8.4	−17.8	−17.7
F1V11PE	−2.1	−2.1	−4.3	−4.2	−5.8	−5.0	−3.4	−2.6
F1V12PE	−3.3	−2.6	−12.6	−12.6	10.3	9.5	14.1	14.2
F1V13PE	2.5	2.4	10.3	10.2	8.3	8.6	6.4	5.5
F2V10PE	−1.6	−2.5	11.8	12.2	—	—	—	—
F2V12PE	6.3	6.2	0.5	0.1	—	—	—	—
F3V10PE	41.9	49.9	32.3	35.4	—	—	—	—
F3V12PE	−39.6	−47.1	−34.9	−37.7	—	—	—	—
F2F1PE	−5.4	−5.7	10.7	10.5	−40.1	−36.5	−19.0	−17.2
F3F2PE	40.6	44.5	14.8	10.4	31.9	31.2	27.3	22.8
N	83	76	31	31	80	70	66	66

	Specification 1		Specification 2		Specification 3		Specification 4	
	Constrained	Unconstrained	Constrained	Unconstrained	Constrained	Unconstrained	Constrained	Unconstrained
			Model 1, $N = 75$, samples with improper solutions					
V1F1PE	−1.6	−2.5	1.1	2.8	2.6	2.6	5.8	5.8
V3F1PE	3.4	2.4	1.7	2.1	5.3	5.3	2.0	2.1
V4F1PE	−18.0	−23.0	35.1	−95.3	38.7	38.5	32.7	33.7
V4F2PE	18.9	24.7	−6.6	52.7	0.1	1.1	30.8	30.3
V6F2PE	50.1	149.2	50.2	28.3	33.6	37.7	—	—
V6F3PE	−105.0	−392.9	−110.4	−50.1	—	—	—	—
V7F2PE	−264.6	−674.5	—	—	—	—	—	—
V7F3PE	118.6	261.8	48.7	50.9	69.2	75.0	66.6	73.9
V9F3PE	−2.9	−1.3	−2.5	−5.9	−9.6	−9.8	−5.4	−5.8
F2F1PE	−9.0	−10.2	−11.9	−9.7	−15.4	−15.9	−2.3	−2.4
F3F2PE	20.8	25.2	11.4	6.0	5.5	3.6	3.8	−0.5
N	38	37	26	24	13	13	12	12

(continued)

TABLE 3 Continued

	Specification 1		*Specification 2*		*Specification 3*		*Specification 4*	
	Constrained	*Unconstrained*	*Constrained*	*Unconstrained*	*Constrained*	*Unconstrained*	*Constrained*	*Unconstrained*
			Model 2, N = 75, samples with improper solutions					
V1F1PE	3.5	3.8	16.0	16.1	2.3	24.1	7.1	6.5
V3F1P	6.3	6.6	3.4	3.4	2.9	21.0	–1.4	–0.9
V4F1PE	15.8	26.2	—	—	12.9	81.0	—	—
V4F2PE	5.3	5.0	11.2	11.3	8.5	38.1	22.0	23.4
V6F2PE	44.0	74.7	19.8	19.6	41.7	136.8	23.9	23.8
V6F3PE	–139.3	–227.2	—	—	–90.0	306.0	—	—
V7F2PE	–175.2	–334.3	—	—	–253.0	839.5	—	—
V7F3PE	80.2	155.9	65.6	72.4	108.7	286.4	65.3	75.1
V9F3PE	1.2	–0.5	–1.9	–1.6	1.6	16.5	4.0	3.0
F1V10PE	0.4	0.8	1.8	1.9	–15.4	31.1	–14.8	–13.0
F1V11PE	0.4	1.1	14.5	14.4	2.0	24.7	–4.4	–5.7
F1V12PE	5.1	5.5	–9.8	–9.8	6.8	32.0	10.8	9.4
F1V13PE	–10.6	–9.8	–10.7	–10.8	0.5	22.7	7.5	7.7
F2V10PE	2.8	1.6	–14.8	–14.8	—	—	—	—
F2V12PE	2.4	2.6	22.8	22.8	—	—	—	—
F3V10PE	32.6	39.7	30.9	33.8	—	—	—	—
F3V12PE	–31.3	–39.4	–38.1	–41.4	—	—	—	—
F2F1PE	–7.5	–9.5	32.4	32.3	–38.4	35.1	–21.5	–22.5
F3F2PE	23.4	29.9	0.2	–3.5	27.9	58.5	9.6	5.0
N	31	28	7	7	47	45	23	24

NOTE: To see the dispersion measures of these parameter estimates (e.g., standard deviation, mean standard error, mean squared error), refer to Tables 1a-1d, 2a-2d, 3a-3d, and 4a-4d in the technical appendix). F1 is η_1, Y1F1 is λ_{11}, F2F1 is β_{21}, and F1X1 is γ_{11}, and so forth.

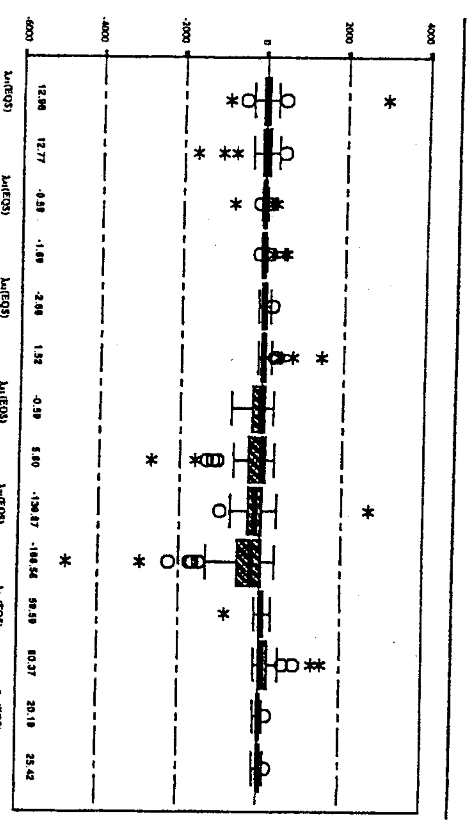

Figure 7: Box Plots of Relative Bias in Selected Parameter Estimates, Comparison of Constrained Versus Unconstrained Results: Model 1, Specification 1, n = 50

under unconstrained estimation, then the estimates from the constrained estimation are likely to be heavily biased. Regardless of the slightly smaller bias that may or may not be gained through using constrained estimation, researchers should be extremely cautious in interpreting estimates of parameters in the presence of an improper solution.

SIGNIFICANCE TESTS

To address our third and final question, we proposed five significance tests that might help to detect whether the cause of improper solution could attribute to sampling fluctuation alone. They were (1) whether the confidence interval (95 percent) includes zero (Van Driel 1978), (2) a z test of the null hypothesis that the error variance is zero versus the alternative that it is smaller than zero, (3) a Wald test, (4) a likelihood ratio chi-square test, and (5) a Lagrangian multiplier test. We applied all of these tests to model 1, specification 1 at both sample sizes 50 and 75. Because we know from the simulation that the model is correctly specified, sampling fluctuation is the only reason why the improper solutions occur.

All five tests yielded the same results. At a sample size of 50, for 86 samples with improper solutions, every confidence interval contained zero, and all z tests, Wald tests, likelihood ratio chi-square tests, and Lagrangian multiplier tests were nonsignificant (using .05 or .10 level). As is shown in the box plots in Figure 8, none of the test statistics was even close to being significant, even for extreme cases (indicated by asterisks in the graph). The results were similar at the sample size of 75. Of 37 samples with improper solutions, the z test, Wald test, and likelihood ratio test were significant for only one sample (results not shown).[12]

These findings are consistent with a previous study, in which only a confidence interval was used (Gerbing and Anderson 1987). However, we remained concerned about the accuracy of the standard errors that underlie the confidence intervals, z test, and Wald test. To examine this, we compared the mean and median of the standard errors for each parameter estimate with a negative error variance in the unconstrained solution to the standard deviation of these negative error variances. We presented the statistics for a few selected error variances that are most likely to have improper solutions in Table 4, based on estimation from model 1, specification 1, $N = 50$. The mean standard errors are bigger than the empirical standard deviations for some estimates (e.g., E1) while smaller for the others (e.g., E7). In fact, there is a negative curvilinear association between the magnitude of the error variances and standard errors (see Figure 9). However, the likelihood ratio test and the Lagrange multiplier test do not make use of standard errors, but these yield the same results as the other tests.

Given that the population error variances are positive, these significance tests result in correct decisions; that is, we should not reject the null hypothesis. As valuable as this is for this model, it leaves open the question of the relative performance of the significance tests when the null hypothesis is true. To explore this question, we generated an additional condition in which the population error variance of E7 was equal to zero, using the correctly specified model 1 for both $N = 50$ and $N = 75$ and repeated these significance tests at an alpha level of 0. For $N = 50$, out of 650 samples, there were 339 samples with improper solutions, 331 (98 percent) of which were associated with E7. The test

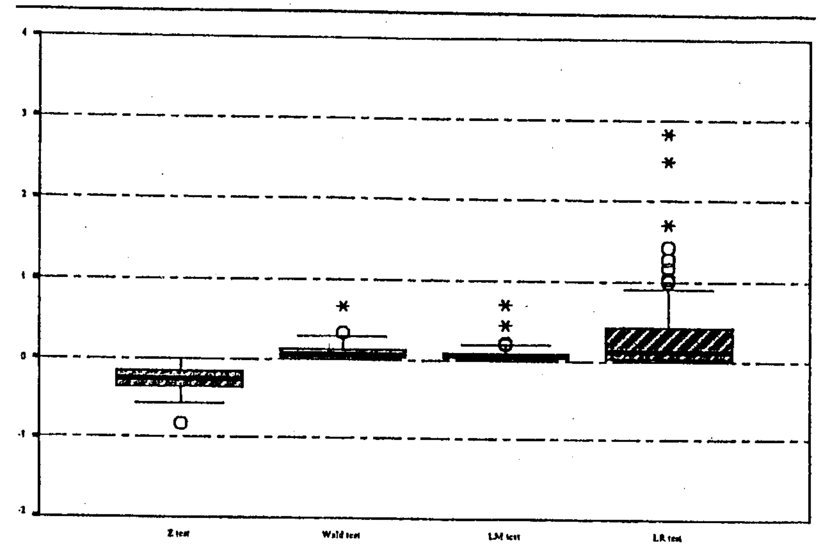

Figure 8: Box Plots of Four Tests of Negative Error Variances: Model 1, Specification 1, $n = 50$ (86 cases)

statistics suggest that the rejection rate of the null hypothesis is much higher (as it should be) since the population value for E7 is now zero. We found that 99.7 percent of the confidence interval includes 0; 0.3 percent of the z tests and Wald tests reject the null; 8.8 percent of the likelihood ratio tests reject zero; none of the Lagrangian multiplier tests is significant. Results from $N = 75$ are quite similar. Given the .05 alpha level for these tests, the confidence intervals, z tests, Wald, and Lagrangian multiplier tests reject less frequently than we would expect. The likelihood ratio test has a higher rejection rate than .05 but is closer to .05 than are the other tests. This higher than nominal rejection rate is consistent with the result from other simulations that find that the chi-square test statistic is too large in smaller samples even under ideal conditions. Despite this, the likelihood ratio test statistic performs the best here and has the desirable property that it can test several error variances simultaneously. However, we would encourage more extensive study and comparisons of this test to the others in the context of testing for improper solutions.

TABLE 4: Estimates of Selected Error Variances and Standard Errors, Model 1, Specification 1, $N = 50$

	Improper Solutions				Proper Solutions			
	N	*Mean*	*Median*	*Standard Deviation*	N	*Mean*	*Median*	*Standard Deviation*
Parameter estimate								
E1	1	–0.065	—	—	632	0.490	0.495	0.156
E4	8	–0.154	–0.154	0.090	625	0.265	0.268	0.105
E6	22	–0.289	–0.119	0.465	611	0.263	0.265	0.097
E7	51	–0.327	–0.108	0.533	582	0.275	0.272	0.115
Standard error								
E1	1	0.335	—	—	632	0.142	0.139	0.030
E4	8	0.692	0.398	0.711	625	0.102	0.097	0.033
E6	22	1.218	0.375	2.709	611	0.098	0.091	0.037
E7	51	0.137	0.129	0.040	582	0.142	0.140	0.030

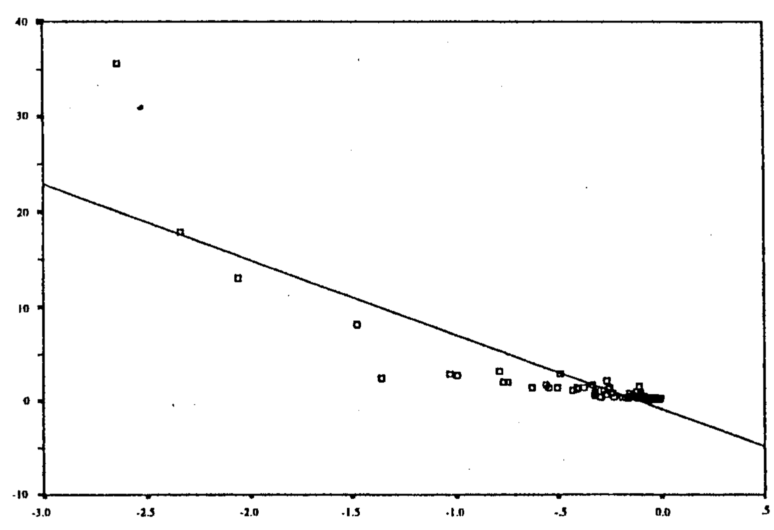

Figure 9: Scattergram of Negative Error Variances and Standard Errors: Model 1, Specification 1, $N = 50$

CONCLUSION

This study investigated improper solutions in the context of two general SEMs and under two sample sizes across several misspecifications. We now return to our three motivating questions for the research: (1) What makes negative error variances more likely to occur? (2) What are the consequences? and (3) What are useful strategies to follow when they occur?

Some of our findings on the contributing factors to improper solutions corroborate earlier research. Improper solutions are more common in small samples than in large ones. This is largely attributable to the greater sampling fluctuations in small samples compared with large ones. However, the relationship between misspecification and improper solutions is more complex than we originally thought. Specifically, we did not find a simple positive relation between measurement model misspecification (omitted paths) and the number of improper solutions. Instead, we must consider the impact of misspecification on the population parameter, the standard deviation, and distribution of the error variance estimator. It is the combination of these factors that determines the probability of improper solutions under misspecification. As our models illustrate, the probability of nonpositive error variance estimates can diminish under misspecification. A clear implication is that researchers should not use negative error variance estimates as an indicator of model misspecification. Of course, it is important to remember that the lack of improper solutions does not necessarily support a model structure either.

It is worth noting that we only considered one type of misspecification in this article: omitted lambda and gamma paths. The consequences of other types of misspecification (e.g., inclusion of nonexistent paths or misspecification of error structure) are not covered and should be included in future works. In addition, we found that none of our misspecified models led to negative error variances in the population. Therefore, shifts in the frequencies of improper solutions were due more to other factors than to misspecification creating a negative error variance in the population.

Our findings on the consequences of negative error variances are a mix of replicating earlier research and revealing new results. Like others, we found no practical difference between samples with only proper solutions and those with improper solutions in terms of chi-square test statistics. We also replicated a prior result that the parameter estimates that were most biased by the improper solutions were those in the locality of the negative variance estimates. Furthermore, the presence of improper solutions tends to mean that the bias in the parameter estimates of the model will be higher than in samples with only proper solutions. This holds across models, specifications, and sample sizes. This finding for our general SEM parallels earlier research using confirmatory factor analyses. As for mean asymptotic standard errors, they are generally quite close to empirical standard deviations for proper solutions. For improper solutions, those parameter estimates that are heavily biased tend to have large mean standard errors and even higher empirical standard deviations.

Some of our other new results address both the consequences and the strategies for handling improper solutions. One tactic that researchers sometimes employ in the presence of improper solutions is constrained estimation. We found that using an inequality constraint sometimes helps convergence in the numerical minimization process. However, those cases that converged under constrained estimation without converging under unconstrained conditions had extremely large negative estimates of variances, which consequently biased the corresponding estimates of lambdas and gammas a great deal. For the samples with improper solutions that achieved convergence under both constrained and unconstrained estimation, the constrained estimates were less biased and had smaller standard errors and mean standard errors. A practical recommendation based on this finding is that researchers should not automatically impose a zero-error variance or an inequality constraint on the error variance. First, they should make sure that the estimator converges even if the error variance is an unrestricted parameter. If not, they risk having severely biased estimators of the model parameters.

It should be strongly noted that even these "less biased" estimates were still more biased than the estimates in samples with only proper solutions. In some ways, this is not that surprising. Consider, for example, if we focused our attention on the samples with the largest positive error variance estimates. We would expect that the parameter estimates from these samples would exhibit more bias than if we took

only those samples that were closer to the center of the distribution of error variances across samples. Our focus on negative error variances is not any different. Examining only those samples with the largest negative error variance estimates highlights the most extreme samples, and we would expect some bias. In the case of negative error variances, we know that these are not right since negative variances are impossible in the population. We cannot similarly detect the most positive error variance values, and they would just be part of the sample of proper solutions.

Based on the evidence from earlier studies and our results, we suggest a preliminary strategy to deal with negative or zero-error variance estimates.[13] Since Heywood estimates might arise from several causes, a strategy must take this into account. With this in mind, our first suggestion is to check the identification of the model. There are rules of identification that can ease the task (e.g., Bollen 1989:88-104, 238-54, 326-33; Davis 1993). If the model is underidentified, the researcher can try to locate the sources of the underidentification and attempt to improve the situation. If the model is identified, then the researcher should estimate the model *without* constraining the error variances to be positive. If the default in an SEM package were to constrain the variance to be nonnegative, then the analyst would have to remove this restriction. The unconstrained models will either converge to a solution or not. If it is nonconvergent, the researcher should check whether this is due to bad starting values or whether the minimization algorithm should be altered.

If the estimation converges, then check to see if there are any negative estimates of error variances. If not, then interpret the results as usual. If there are negative error variance estimates, then determine whether these are due to outliers or influential cases (Arbuckle 1997; Bollen 1987; Bollen and Arminger 1991; Cadigan, 1995). In the event that the negative error variances are due to influential cases, then investigate the causes of these unusual values and explore ways to correct them. Alternatively, with no influential cases, the researcher should screen for empirical underidentification (see Kenny 1979; Rindskopf 1984). Assuming that the negative error variance estimates are not due to influential cases or empirical underidentification, the next step is to test whether the negative error variance estimates might be due to sampling fluctuations. We have recommended five possible significance tests (confidence interval, z test, Wald test, chi-square ratio test, and Lagrange multiplier's test). The latter three have the advantage that they can simultaneously test two or more error variances. If the significance tests suggest that the error variance(s) are below zero, then the researcher should suspect model misspecification. This is true since a researcher who follows our sequence of steps will have already ruled out the most likely other causes.

A nonsignificant negative error variance estimate is consistent with the idea of sampling variability leading to the negative estimates. Our results suggest that the next step should be to reestimate the model constraining the error variances to zero or a small positive number. This could reduce the bias in the parameter estimates. Of course, this is given that the magnitude of the negative error variances is not big. If the estimate were very big in magnitude, imposing a constraint at zero would not be helpful since it might still have a severe bias in the parameter estimates related to the error variance. If the negative estimate of variance is not far from zero, then the constraint at zero may provide less biased parameter estimates than no constraint. The bigger the negative estimates, the less helpful it is to impose the constraint. If the population value of the variance were known or could be estimated, then the constraint should be set to the population value, although this rarely happens in practice. In sum, the resulting estimates after the constraint should be interpreted as usual except that those parameters most closely associated with the troublesome error variance estimates are likely to have higher bias than estimates that result in a fully proper solution without the constraints. For example, the factor loading to the indicator with the original negative error variance is likely to be too big in absolute magnitude, while the factor loadings for the other indicators influenced by the same factor might be too small.

We regard the preceding analysis strategy as preliminary. We say this because of the limitations of our study and the other literature in this area. First, there is an inherent limitation in Monte Carlo studies as to the representativeness of the model structures researchers consider. It is difficult to know whether results from one set of models will generalize to the models typically estimated in practice. Although we have departed from prior research in including both correctly and incorrectly specified models and factor analysis as well as latent variable structures, it is impossible to be certain of the limits of the findings. A related restriction is that research, including ours, has examined data that were simulated from a multivariate normal distribution. A natural question is, How sensitive are our findings to having variables from nonnormal distributions? However, the literature on the robustness conditions for these models and nonnormal data (e.g., Satorra 1990), the corrections to the standard errors and test statistics for nonnormality, and the possibility of bootstrapping for significance testing and confidence intervals suggest that the nonnormality is something that might not be as serious as other problems.

A more complicated issue is the relation between misspecification and the occurrence of negative error variances. Our simulation results demonstrate that the relation is more complex than previously thought. Future research designs should seek to more fully determine the conditions under which misspecification leads to greater or fewer numbers of negative estimates of error variances. Although part of the answer should lie in the magnitude of the population error variances in correctly and incorrectly specified models, our findings suggest that this is not the entire answer.

Finally, Van Driel's (1978) confidence intervals and the tests of significance that we propose bear closer scrutiny. It would be useful to comparatively evaluate each test statistic to see which performs most accurately as a test for negative error variances. This is a key hypothesis to test since the outcome of the test points us toward sampling fluctuations or the more serious issue of misspecification depending on what we find. Our research has provided fresh evidence about improper solutions, but it also raises many new issues for investigation.

NOTES

1. See Dijkstra (1992) for theoretical results relevant to comparing the estimates with and without the constraint.

2. We realize that there are alternative ways in which the degree of misspecification might be measured (e.g., the mean bias in parameter estimates could gauge the amount of misspecification error). However, we stay with the magnitude of the fitting function evaluated at the population covariance matrix as our measure of the degree of misspecification because in addition to providing a metric for misspecification, it also is directly tied to the power of the chi-square test statistic (Satorra and Saris 1985).

3. These were identified by a condition code, "constrained at the lower bound" in EQS, with at least one of the error variance estimates fixed to zero.

4. A reviewer raised the possibility that empirical underidentification could be related to the negative error variances in some of the models. Although a possibility, we found no evidence to suggest that this is an issue for our models.

5. Detailed information on all parameter estimates across all conditions (models, sample sizes, and estimation techniques) appears in the technical appendix. Almost unequivocally across conditions, we find that excluding the improper solutions leads to a decrease in parameter bias.

6. There are three exceptions: λ_{11}, λ_{31}, and λ_{32}. For the estimates of these three parameters, the samples with improper solutions tend to have lower parameter bias. This is discussed in greater detail below.

7. For the parameter estimates that are severely biased, the standard deviation, standard errors, and mean squared errors are also bigger (these results appear in the technical appendix).

8. We focus on results with $N = 50$ since a similar pattern is observed with $N = 75$.

9. Numbers of improper solutions under constrained estimation are slightly larger than unconstrained estimation, as reported in Table 3, since imposing constraints helps achieve convergence. The comparison does not look different when those nonconverged cases under unconstrained estimation were excluded. Box plots presented in Figure 7 as well as those in the technical appendix are based on improper solutions that achieved convergence under both types of estimation.

10. These estimates are additionally conservative because they include only samples that converged under both types of estimation. As will be discussed below, the samples that would not converge under unconstrained estimation were especially likely to show bias.

11. Box plots for all specifications and sample sizes are available in the technical appendix.

12. That was an extreme case. The error variance is –11.74, and the standard error is 3.72.

13. See Kano (1998), Sato (1987), and Van Driel (1978) for other discussions of strategies to handle improper solutions.

REFERENCES

Anderson, James C. and David Gerbing. 1984. "The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness-of-Fit Indices for Maximum Likelihood Confirmatory Factor Analysis." *Psychometrika* 49:155-73.

Arbuckle, James. 1997. *AMOS Users' Guide: Version 3.6*. Chicago: SmallWaters.

Bentler, Peter M. 1997. *EQS: Structural Equations Program Manual*. Los Angeles: BMDP Statistical Software.

Bollen, Kenneth A. 1987. "Outliers and Improper Solutions: A Confirmatory Factor Analysis Example." *Sociological Methods & Research* 15:375-84.

———. 1989. *Structural Equations With Latent Variables*. New York: John Wiley.

Bollen, Kenneth A. and Gerhard Arminger. 1991. "Observational Residuals in Factor Analysis and Structural Equation Models." Pp. 235-62 in *Sociological Methodology*, edited by P. V. Marsden. Washington, DC: American Sociological Association.

Boomsma, Anne. 1983. *On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Nonnormality*. Amsterdam, the Netherlands: Sociometric Research Foundation.

———. 1985. "Nonconvergence, Improper Solutions, and Starting Values in LISREL Maximum Likelihood Estimation." *Psychometrika* 50:229-42.

Cadigan, N. G. 1995. "Local Influence in Structural Equation Models." *Structural Equation Modeling* 2:13-30.

Davis, Walter R. 1993. "The FC1 Rule of Identification for Confirmatory Factor Analysis: A General Sufficient Condition." *Sociological Methods & Research* 21:403-37.

Dijkstra, Taeke K. 1992. "On Statistical Inference With Parameter Estimates on the Boundary of the Parameter Space." *British Journal of Mathematical and Statistical Psychology* 45:289-309.

Dillon, William R., Ajith Kumar, and Narendra Mulani. 1987. "Offending Estimates in Covariance Structure Analysis: Comments on the Causes and Solutions to Heywood Cases." *Psychological Bulletin* 101:126-35.

Gerbing, David W. and James C. Anderson. 1987. "Improper Solutions in the Analysis of Covariance Structures: Their Interpretability and a Comparison of Alternate Respecifications." *Psychometrika* 52:99-111.

Joreskog, K. and D. Sorbom. 1993. *LISREL 8: Structural Equation Modeling With SIMPLIS Command Language*. Hillsdale, NJ: Lawrence Erlbaum.

Joreskog, Karl G. and D. N. Lawley. 1968. "New Methods in Maximum Likelihood Factor Analysis." *British Journal of Mathematical and Statistical Psychology* 21:85-96.

Kano, Yutaka. 1998. "Improper Solutions in Exploratory Factor Analysis: Causes and Treatments." Pp. 375-82 in *Advances in Data Science and Classification: Proceedings*, edited by A. Rizzi. Berlin: Springer.

Kenny, David. 1979. *Correlation and Causality*. New York: John Wiley.

Luijben, Thom. 1989. *Statistical Guidance for Model Modification in Covariance Structure Analysis*. Amsterdam, the Netherlands: Sociometric Research Foundation.

Paxton, Pamela, Kenneth Bollen, Patrick Curran, James Kirby, and Feinian Chen. Forthcoming. "Monte Carlo Experiments: Design and Implementation." *Structural Equation Modeling*.

Rindskopf, David. 1984. "Structural Equation Models: Empirical Identification, Heywood Cases, and Related Problems." *Sociological Methods & Research* 13:109-19.

Sato, Manabu. 1987. "Pragmatic Treatment of Improper Solutions in Factor Analysis." *Annals of the Institute of Statistics and Mathematics Part B* 39:443-55.

Satorra, Albert. 1990. "Robustness Issues in Structural Equation Modeling: A Review of Recent Developments." *Quality & Quantity* 24:367-86.

Satorra, Albert and Willem Saris. 1985. "The Power of the Likelihood Ratio Test in Covariance Structure Analysis." *Psychometrika* 50:83-90.

Van Driel, Otto P. 1978. "On Various Causes of Improper Solutions in Maximum Likelihood Factor Analysis." *Psychometrika* 43:225-43.

Feinian Chen is a Ph.D. candidate in the Department of Sociology at the University of North Carolina at Chapel Hill and a trainee at the Carolina Population Center. Her research interests include family, demography, and quantitative methods. Her dissertation examines the relationship between socioeconomic change and changes in family structures and familial relationships in China and Russia.

Kenneth A. Bollen is the director of the Odum Institute for Research in Social Science, the Zachary Smith Professor of Sociology, and a fellow of the Carolina Population Center at the University of North Carolina at Chapel Hill. His primary area of statistical research

is structural equation modeling (SEM). He is applying SEMs in the analysis of the relation of socioeconomic status to family planning and child health in developing countries. Other research projects in progress are on comparative liberal democracy and biases in ratings of graduate school departments. He is the author of Structural Equations With Latent Variables *(John Wiley) and coeditor (with J. S. Long) of* Testing Structural Equation Models *(Sage). Recent articles in these topic areas have been published in* American Sociological Review, Psychological Methods, Comparative Political Studies, Demography, *and* Structural Equation Modeling.

Pamela Paxton is assistant professor of sociology at the Ohio State University. Her research interests include political sociology, quantitative methods, and stratification. Some of her recent publications have appeared in the American Journal of Sociology, Comparative Political Studies, *and the* American Sociological Review. *She is currently investigating the relationship between social capital and democracy cross nationally.*

Patrick J. Curran is an assistant professor in the Department of Psychology at the University of North Carolina at Chapel Hill. His primary area of research focuses on the development of new quantitative methods for longitudinal data analysis and the application of these methods to study aggression and substance use in children and adolescents. He is currently receiving funding from the National Institute of Drug Abuse to further support this work. He has recently published articles in Psychological Methods, Structural Equation Modeling, Journal of Consulting and Clinical Psychology, *and* Psychological Assessment.

James B. Kirby is a service fellow at the Agency for Healthcare Research and Quality. His research focuses on how changes in family structure influence children's access to health care and their health-related behaviors.

Unit nonresponse in the second wave of a panel survey is related to several respondent characteristics and to the interviewer of that wave. More striking is the effect of the interviewer of the first wave, who is not involved in the second interview. To analyze both interviewer effects simultaneously, the authors use a multilevel cross-classified model. In that analysis, the effect of the interviewer of the second wave almost disappears. That effect turns out to be at least partly spurious due to a correlation of both interviewer effects. The authors conclude that the interviewer of the first interview is very important regarding participation in the subsequent waves of a panel survey.

The Effects of Interviewer and Respondent Characteristics on Response Behavior in Panel Surveys

A Multilevel Approach

JAN PICKERY
GEERT LOOSVELDT
ANN CARTON
University of Leuven, Belgium

In this article, we analyze the nonresponse in the second wave of a panel survey. We restrict ourselves to the respondents who were interviewed during the first wave of the panel but refused to cooperate again in the second wave. We look for interviewer effects on that refusal rate. In particular, we control for the interviewer of the first wave who has an impact on the respondent's experience with the interview and consequently can affect his or her decision to cooperate again.

Among other things, refusals make up an important component of nonresponse. A refusal is an active act of the respondent and can be considered as a crucial aspect of respondent behavior. Several models try to explain this kind of respondent behavior (see, e.g., Goyder 1987;

AUTHORS' NOTE: *We would like to thank the anonymous reviewers for their comments that certainly helped us to clarify our arguments.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 29 No. 4, May 2001 509-523