



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing

Rajdeep Grewal, Joseph A. Cote, Hans Baumgartner,

To cite this article:

Rajdeep Grewal, Joseph A. Cote, Hans Baumgartner, (2004) Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing. Marketing Science 23(4):519-529. <https://doi.org/10.1287/mksc.1040.0070>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

© 2004 INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing

Rajdeep Grewal

Smeal College of Business Administration, Pennsylvania State University,
University Park, Pennsylvania 16802-3007, rug2@psu.edu

Joseph A. Cote

Washington State University, 14204 Salmon Creek Avenue, Vancouver, Washington 98686, cote@vancouver.wsu.edu

Hans Baumgartner

Smeal College of Business Administration, Pennsylvania State University,
University Park, Pennsylvania 16802-3007, jxb14@psu.edu

The literature on structural equation models is unclear on whether and when multicollinearity may pose problems in theory testing (Type II errors). Two Monte Carlo simulation experiments show that multicollinearity can cause problems under certain conditions, specifically: (1) when multicollinearity is extreme, Type II error rates are generally unacceptably high (over 80%), (2) when multicollinearity is between 0.6 and 0.8, Type II error rates can be substantial (greater than 50% and frequently above 80%) if composite reliability is weak, explained variance (R^2) is low, and sample size is relatively small. However, as reliability improves (0.80 or higher), explained variance R^2 reaches 0.75, and sample becomes relatively large, Type II error rates become negligible. (3) When multicollinearity is between 0.4 and 0.5, Type II error rates tend to be quite small, except when reliability is weak, R^2 is low, and sample size is small, in which case error rates can still be high (greater than 50%). Methods for detecting and correcting multicollinearity are briefly discussed. However, since multicollinearity is difficult to manage after the fact, researchers should avoid problems by carefully managing the factors known to mitigate multicollinearity problems (particularly measurement error).

Key words: multicollinearity; measurement error; structural equation models

History: This paper was received June 26, 2002, and was with the author 12 months for 2 revisions; processed by Michel Wedel.

1. Introduction

Perhaps due to its ability to account for measurement error and manage multiple endogenous constructs, structural equation modeling (SEM) has become a commonly used tool for theory testing (Steenkamp and Baumgartner 2000). Important methodological advances, such as methods for treating heterogeneity (Ansari et al. 2000, Jedidi et al. 1997), have broadened the scope of the technique, and even marketing practitioners are adopting it in growing numbers (Grapentine 2000). The widespread use (and misuse) of SEM has led to a growing literature that addresses various application problems (Baumgartner and Homburg 1996, Hulland et al. 1996). One of these problems is multicollinearity, i.e., high correlations among the latent exogenous constructs. Unfortunately, little is known about the effects of multicollinearity in SEM, and researchers frequently pay little attention to it. For example, 42 articles using either confirmatory factor analysis (CFA) or SEM were published in the 1999 and 2000 issues of the *Journal of Marketing*, *Journal of Marketing Research*, and *Journal*

of the Academy of Marketing Science.¹ Potential multicollinearity problems could be assessed for 31 of the studies through the published correlation matrix. Nine studies (29%) reported high correlations ($\rho = 0.75$ to 0.95) among latent constructs. Despite these high correlations, not a single article discussed how multicollinearity might have affected the results. Such oversight could be critical because managing multicollinearity is a growing concern in applied research and theory development (Farley et al. 1998).

The dismissal of multicollinearity problems appears to have two basic causes. First, some ambiguity exists about whether multicollinearity is really a problem in SEM. Many researchers seem to think that structural equation models are robust against multicollinearity (Malhotra et al. 1999), with some going so far as to explicitly state that SEM can remedy multicollinearity problems. For example, Verbeke and Bagozzi (2000, p. 93) used LISREL to “avoid problems of

¹ During this period *Marketing Science* had only one article that used SEM. Thus, we decided not to include *Marketing Science* in the survey.

multicollinearity.” Similarly, Maruyama (1998, p. 61) argues that “structural equation approaches can help deal with some cases where the correlations among predictors are large.” One reason for this optimism seems to be that if highly correlated variables can be regarded as indicators of a common underlying construct, multicollinearity problems can be avoided. In contrast, some researchers have warned that multicollinearity can lead to SEM estimates far from the true parameters and large standard errors of the estimates (Jagpal 1982, Grapentine 2000).

Second, researchers may ignore multicollinearity because of practical considerations. Existing guidelines about when multicollinearity is likely to cause problems are often ambiguous, procedures for mitigating multicollinearity are frequently of limited usefulness and, most importantly, little is known about how to deal with multicollinearity in the context of SEM. The best solution would be to avoid multicollinearity problems in the first place. However, as Rindskopf (1984, p. 111) notes, “[t]he exact values that might cause or prevent the problems discussed are rather fuzzy.” Since the conditions that cause multicollinearity problems are unclear, it is difficult to avoid these conditions. Thus, researchers easily dismiss messy multicollinearity problems by assuming SEM is robust, or ignore the problems because little is known about multicollinearity.²

The disregard for multicollinearity problems raises important questions. Is it true, as some claim, that SEM can be used to deal with multicollinearity? If not, then how can we minimize the effects of multicollinearity? These uncertainties led Kaplan (1994, p. 211) to suggest that “examin[ing] the effects of increasing multicollinearity on [SEM] parameter estimates and standard errors . . . is an important area for future research.” To illuminate these issues, we report two Monte Carlo simulation experiments in which we vary the level of multicollinearity, measurement error, amount of explained variance, relative importance of the exogenous variables, and sample size. We study the effects of multicollinearity on the accuracy of coefficient and standard-error estimates, as well as inferences errors. Our research takes important steps toward identifying conditions under which

multicollinearity and measurement error in SEM are likely to cause misleading tests of theory.

2. Background

The literature on multicollinearity can be divided into three major topics: (1) under what conditions multicollinearity will occur; (2) how multicollinearity problems can be detected; and (3) how multicollinearity should be managed. Our review of the literature will show that we know relatively little about the conditions that lead to multicollinearity problems in SEM. Although we do have tools for detecting when multicollinearity may be affecting estimates, these techniques are often ambiguous. Lastly, there are some remedial actions that can be taken when multicollinearity exists, but they may be difficult to implement, and in general the evidence regarding their practical effectiveness is limited.

2.1. When Does Multicollinearity Pose a Problem in SEM?

Mason and Perreault (1991) have documented the conditions under which multicollinearity may pose problems in regression. They show that multicollinearity leads to inaccurate estimates of coefficients and standard errors as well as inference errors, but they also argue that the problem should not be viewed in isolation, and that a high R^2 and large sample size can offset the problems caused by multicollinearity. Although these factors should also be relevant in the context of SEM, their work is silent about certain issues that are specific to SEM, most notably, measurement error. The ability of SEM to incorporate measurement error makes it difficult to assess the impact of multicollinearity on parameter estimates (Bollen 1989). Mason and Perreault (1991) show that increasing explained variance in the dependent variable mitigates the effects of multicollinearity. Removing measurement error should increase the amount of variance explained by the structural model and, by extension, mitigate multicollinearity. However, measurement error also attenuates correlations among the exogenous variables. The presence of measurement error is likely to mask the true correlation among latent exogenous constructs. Thus, controlling for measurement error should result in higher correlations among the exogenous constructs than not controlling for measurement error.³ As Maruyama (1998, p. 61) points out, “although latent variable approaches help in most instances by removing measurement and specification error from variables, they ironically may make high multicollinearity appear in cases where it

² Theoretically, as Rindskopf (1984, p. 115) notes, “the problem of multicollinearity can be viewed as another [special] case of empirical underidentification” (also see Kenny 1979). When models are empirically underidentified, estimates are likely to be unstable and may not accurately reflect population values. Unfortunately, it is unclear at what point empirical underidentification will become a problem. As Rindskopf (1984, p. 111) states, “The exact values that might cause or prevent the problems discussed [empirical underidentification] are rather fuzzy.” All we know is that once multicollinearity reaches a critical point, the structural equation model will become underidentified and parameter estimates may be problematic.

³ It should be noted that removing measurement error can sometimes reduce correlations (for examples of such cases see Cote and Buckley 1988, Fornell and Larcker 1981b).

previously has not been a problem” (also see Fornell and Larcker 1981b, Heise 1969).⁴ Thus, it is unclear if eliminating measurement error will mitigate multicollinearity due to the beneficial effect of an increase in R^2 or exacerbate it by increasing the correlations among the exogenous constructs.

2.2. Detecting Multicollinearity

Various tools have been proposed to diagnose situations in which multicollinearity among the predictors may create estimation and inference problems, although little is known about the usefulness of these techniques for SEM. Kaplan (1994) offers one of the few discussions of conditioning problems (of which multicollinearity is a special case) for SEM. He reviews several general methods that can be used to detect multicollinearity, including inspection of the (1) correlation matrix of the predictor variables, (2) correlation matrix of the path coefficients, (3) determinant of the correlation matrix of the predictor variables, (4) sign of the path coefficients, and (5) variance inflation factors. However, he calls these methods “more or less ad hoc” (Kaplan 1994, p. 201). Kaplan also proposes a general conditioning diagnostic for structural equation models based on the condition number of the correlation matrix of the estimated parameters. Although the procedure is useful for detecting highly collinear parameter estimates, it may not be able to diagnose multicollinearity among the observed variables. Further, as condition indices are model specific, rules of thumb about what constitutes a large condition number may be misleading (Bollen 1989).

Schmidt and Muller (1978) also discuss the problem of multicollinearity in the context of SEM and argue that in multistage models typical of many applications of SEM, commonly used multicollinearity diagnostics should be applied “at every stage of a given ‘causal model’” (p. 267). They discuss three specific methods for assessing multicollinearity: (1) the multiple correlation between the independent variables; (2) the Haitovsky test, which assesses singularity of the correlation matrix of the independent variables; and (3) the determinant and eigenvalues of the correlation matrix of the independent variables. However, these methods have limitations because they do not provide clear guidelines about when multicollinearity is likely to cause inference errors.

2.3. Managing Multicollinearity

When one of the telltale signs of multicollinearity occurs (nonsignificant coefficient estimates even though the overall regression is highly significant, “wrong” signs of the coefficients, unstable parameter

estimates), researchers generally feel compelled to do something about the problem. Unfortunately, many of the solutions are ad hoc (e.g., dropping a variable, which may lead to a misspecification error, or restricting parameters, which is not always possible) or impractical (e.g., obtaining additional data that do not suffer from multicollinearity). Among the more formal remedies is ridge estimation, which introduces a small amount of bias in return for greater efficiency. Jagpal (1982) developed a partial-least-squares (PLS)-based ridge estimator to deal with multicollinearity in structural equation models (also see Schmidt and Muller 1978, Maruyama 1998). Although ridge estimation has been incorporated into SEM programs such as LISREL to deal with cases in which the variance-covariance matrix is not positive definite, little is known about the practical benefits of using ridge estimation in SEM contexts (Kennedy 1992). Alternatively, it is possible to move away from SEM and into the realm of regression by forming composite indicators (e.g., factor scores). In this case, various other estimation techniques, such as the equity estimator (Krishnamurthi and Rangaswamy 1987), are available. However, a serious problem with using composite indicators is that one of the main advantages of using SEM, the ability to explicitly account for measurement error, is lost. In general, none of the approaches for managing multicollinearity seems entirely satisfactory.

2.4. The Need to Identify Conditions That Lead to Multicollinearity Problems

In summary, although we have tools that can identify situations in which multicollinearity may cause problems, there are two issues with existing approaches. First, many of the guidelines are ambiguous (e.g., at what point is a correlation between independent variables too high) and potentially misleading (e.g., if condition numbers are model dependent, absolute rules of thumb are meaningless). Second, the major conclusion of Mason and Perreault (1991) was that the “deleterious effect of a given level of multicollinearity should be viewed in conjunction with other factors known to affect estimation accuracy” (p. 268). Guidelines based solely on the amount of shared variance among predictor variables ignore this recommendation. Given that multicollinearity is difficult to mitigate once it is present in data, the best strategy would be to avoid the problem in the first place. Knowing the conditions likely to lead to multicollinearity problems and the factors that help offset its damaging effects can help researchers develop strategies to mitigate its effects.

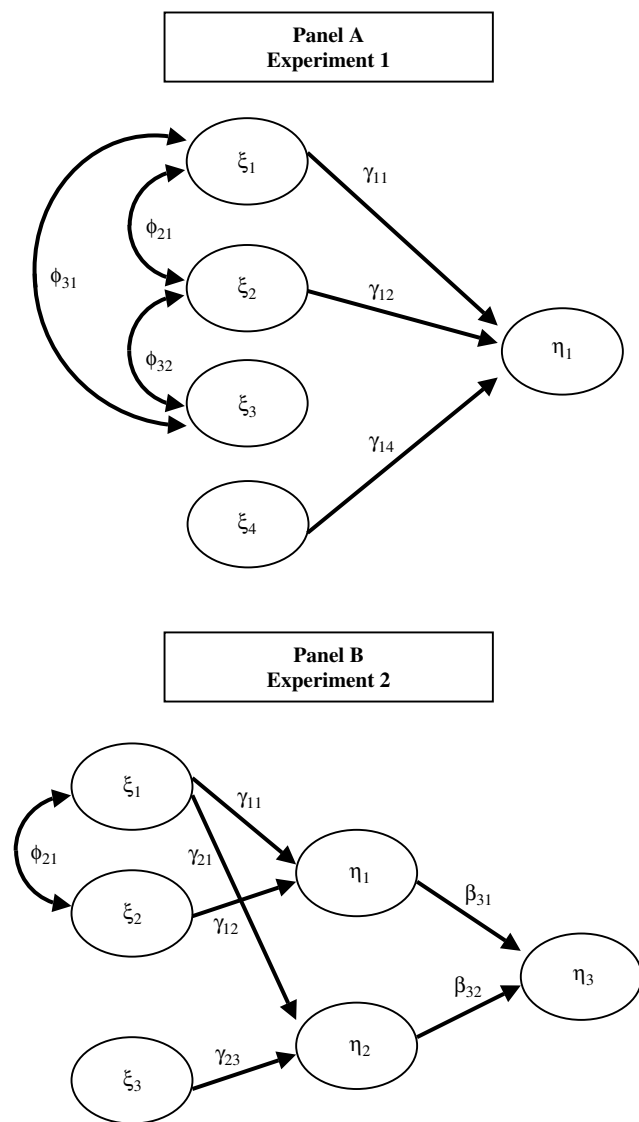
3. Monte Carlo Simulations

To capture situations typically encountered by marketing researchers, Experiment 1 examines a scenario

⁴ For example, Grapentine (2000) found that a correlation of 0.66 between two constructs increased to 0.95 after correcting for measurement error.

with four latent exogenous variables and one latent endogenous variable, with four items for each latent construct (Bagozzi and Baumgartner 1994, Panel A, Figure 1). The levels of the design factors (i.e., multicollinearity, measurement error, R^2 , relative weights of the exogenous variables, and sample size) were manipulated to reflect conditions typically encountered by marketing researchers (Mason and Perreault 1991, Baumgartner and Homburg 1996). Since marketing researchers often deal with multiple endogenous variables and since SEM is particularly suited for this type of analysis, Experiment 2 extends Experiment 1 to the case of multiple endogenous variables (Panel B, Figure 1).

Figure 1 Models for the Two Experiments*



* In both experiments each latent construct is measured by four items. For clarity, these item details are not depicted in the figure. Furthermore, we only depict relationships between latent constructs that were specified to be nonzero (positive). Similarly, we only depict nonzero correlations.

Based on the design factors and the specific values chosen for the other model components, we formulated various population covariance matrices and input these matrices into EQS version 5.7b to simulate our data (Bentler and Wu 1995). Following Anderson and Gerbing (1984) and Paxton et al. (2001), we generated 200 proper solutions for each cell of the design. A solution was regarded as proper if (1) there were no convergence problems, (2) the covariance matrix of the latent variables was positive definite, and (3) there were no linear dependencies among the estimated parameters.⁵

3.1. Design Factors

In Experiment 1 we replicated Mason's and Perreault's (1991) multicollinearity ranges at four levels of 0.2 to 0.5, 0.4 to 0.65, 0.6 to 0.8, and 0.8 to 0.95. For example, at the first level of multicollinearity ξ_3 has a correlation of 0.2 with ξ_1 and ξ_2 , and ξ_1 and ξ_2 are correlated at 0.5. In accordance with acceptable levels reported in the literature (Cote and Buckley 1987), we systematically varied measurement error at three levels. Specifically, the composite reliabilities of ξ_1 through ξ_4 and η_1 were all set at 0.7, 0.8, or 0.9, with uniform loadings for all indicators on a given construct. Following Mason and Perreault, we also selected three levels of total explained variance in the endogenous construct ($R^2 = 0.25, 0.50$, or 0.75) and two sets of weights for the relative impact of the four exogenous constructs (2, 1, 0, 1 and 2, 2, 0, 1 for γ_{11} , γ_{12} , γ_{13} , and γ_{14} , respectively; Panel A of Figure 1). Specifically, the impact of ξ_1 on η_1 was twice as large as the impact of ξ_2 on η_1 in the first case, but equal in the second case. Finally, we set sample size at two levels, using ratios of 3:1 and 6:1 for the number of observations to the number of parameters estimated (Baumgartner and Homburg 1996). Therefore, we have four levels of multicollinearity, three levels of measure reliability and R^2 , and two levels of coefficient weights and sample size, giving us 144 different combinations of the design factors. We generated 200 proper solutions for each cell, resulting in 28,800 sets of estimates.

In Experiment 2, multicollinearity was varied at three levels by systematically altering the correlation between ξ_1 and ξ_2 ($\phi_{21} = 0.4, 0.6$, or 0.8 ; Panel B of Figure 1). We also varied composite reliability at four levels (0.6, 0.7, 0.8, and 0.9), R^2 at three levels (0.25, 0.50, or 0.75—same for all endogenous variables), and sample size at two levels (i.e., ratios of 3:1 and 6:1 for the number of observations to the number of parameters estimated). Thus, we have three levels of multicollinearity and R^2 , four levels of measure reliability, and two levels of sample size, giving us 72 different

⁵ A technical appendix, available either from the authors on request or the *Marketing Science* website, details the data-generation framework.

Table 1 Analysis of Variance Results for Estimates of Path Coefficients and Standard Errors

Experiment 1																
Sources of variance	% Variance explained in coefficient estimate				% Variance explained in estimate of standard error											
	γ_{11}	γ_{12}	γ_{13}	γ_{14}	γ_{11}	γ_{12}	γ_{13}	γ_{14}								
Overall model	32.07 ^a	33.80 ^a	21.35 ^a	12.39 ^a	71.31 ^a	71.36 ^a	53.48 ^a	40.96 ^a								
Multicollinearity	22.38 ^a	24.21 ^a	8.61 ^a	0.39 ^a	52.01 ^a	52.37 ^a	24.83 ^a	8.46 ^a								
Reliability	3.53 ^a	2.85 ^a	3.83 ^a	2.83 ^a	11.77 ^a	10.37 ^a	14.00 ^a	17.92 ^a								
R-Square	4.54 ^a	5.86 ^a	6.90 ^a	6.70 ^a	2.41 ^a	2.27 ^a	4.03 ^a	4.47 ^a								
Coefficient weights	0.00	0.00	0.02 ^d	0.02 ^d	0.01 ^b	0.00	0.01	0.05 ^a								
Sample size	1.11 ^a	0.52 ^a	1.64 ^a	2.08 ^a	2.84 ^a	4.32 ^a	8.31 ^a	6.98 ^a								
Experiment 2																
Sources of variance	% Variance explained in coefficient estimate								% Variance explained in estimate of standard error							
	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}	β_{13}	β_{23}	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}	β_{13}	β_{23}
Overall model	20.05 ^a	20.66 ^a	16.56 ^a	15.61 ^a	16.70 ^a	12.38 ^a	13.10 ^a	8.96 ^a	83.11 ^a	83.77 ^a	85.53 ^a	70.62 ^a	73.79 ^a	80.67 ^a	82.96 ^a	83.81 ^a
Multicollinearity	5.41 ^a	6.48 ^a	0.19 ^a	2.77 ^a	4.04 ^a	0.02	0.00	0.08 ^b	24.82 ^a	25.80 ^a	1.56 ^a	12.18 ^a	16.02 ^a	0.12 ^a	0.03 ^a	0.67 ^a
Reliability	7.30 ^a	6.37 ^a	4.37 ^a	7.90 ^a	7.99 ^a	7.07 ^a	5.20 ^a	4.30 ^a	27.08 ^a	26.80 ^a	25.23 ^a	40.00 ^a	36.25 ^a	45.84 ^a	44.85 ^a	41.80 ^a
R-Square	5.11 ^a	5.56 ^a	8.26 ^a	2.03 ^a	2.15 ^a	2.46 ^a	2.23 ^a	2.19 ^a	21.27 ^a	21.26 ^a	43.40 ^a	5.47 ^a	8.26 ^a	18.25 ^a	19.10 ^a	20.88 ^a
Sample size	1.92 ^a	2.00 ^a	3.43 ^a	2.25 ^a	2.08 ^a	2.64 ^a	5.52 ^a	2.23 ^a	8.87 ^a	8.80 ^a	14.46 ^a	9.97 ^a	10.04 ^a	14.91 ^a	18.17 ^a	19.50 ^a

Note. Only main effects are shown for simplicity.

^a $p < 0.001$ (F-test).

^b $p < 0.005$ (F-test).

^c $p < 0.010$ (F-test).

^d $p < 0.050$ (F-test).

combinations of the design factors. With 200 replications in each cell, there are 14,400 sets of estimates.

4. Simulation Results

4.1. Experiment 1

We had to generate a total of 30,322 samples (an additional 5%) in order to obtain 200 proper solutions in each cell of the design. Most of the 1,522 improper solutions (96%) occurred in the highest multicollinearity condition ($\rho_{\xi_1, \xi_2} = 0.95$), and the incidence of improper solutions seems to be nonnegligible in these cases. For example, in the condition in which correlations ranged from 0.8 to 0.95, composite reliability was 0.7, sample size was 3:1, R^2 was 0.75, and γ_1 and γ_2 were unequal in magnitude, 44% of the solutions were improper. Thus, one difference relative to the findings of Mason and Perreault (1991) is that very high levels of multicollinearity are likely to result in improper solutions, especially when composite reliability is low and sample size is small. Because we only consider proper solutions in the sequel, the findings are conditional on having obtained a proper solution.

To assess the efficacy of the design variables, we computed the accuracy of the estimated coefficients and their standard errors as the absolute difference between an estimate and its true value.⁶ The design

variables explained 32%, 34%, 21%, and 12% of the variance in the accuracy of the estimated coefficients of γ_{11} , γ_{12} , γ_{13} , and γ_{14} , respectively (see Table 1). Similar to Mason and Perreault (1991), the explained variance for the accuracy of the estimated standard errors was much higher at 71%, 71%, 53%, and 41% for γ_{11} , γ_{12} , γ_{13} , and γ_{14} , respectively. Although several higher-order effects were significant because of the very large sample size, the four main effects of degree of multicollinearity, measure reliability, R^2 , and sample size generally accounted for most of the explained variance in each dependent variable (92% for the standard error of γ_{14} and over 96% for the other parameters). Coefficient weights did not explain meaningful amounts of variance in the inaccuracy of coefficients and their standard errors. The adverse effects of multicollinearity were most pronounced for γ_{11} and γ_{12} , which was expected because multicollinearity was manipulated by varying the correlation between ξ_1 and ξ_2 . Over 70% of the explained variance in the inaccuracy of both the coefficients and standard errors was accounted for by multicollinearity in the case of γ_{11} and γ_{12} . Measure reliability, R^2 , and, to a lesser degree, sample size also explained substantial portions of the variance in the inaccuracy of coefficients and standard errors. Reliability was a particularly important influence in the case of the standard errors of the coefficients.

As theory testing usually involves ascertaining the direction (positive or negative) and significance of a parameter estimate, researchers are generally most

⁶ As the data for accuracy of the coefficients and standard errors were not normal, we applied a natural logarithm transformation before conducting the analysis of variance.

concerned about inference errors, specifically, Type II errors (i.e., failures to detect a significant effect). Thus, one needs to look at the results for the coefficients and standard errors in tandem. A high standard error can result in a Type II error even when the coefficient estimate is correct, as can a wrong coefficient estimate even when the standard error is correct. The percentage of Type II errors in each cell of the design is shown in Table 2. It is apparent that Type II error rates vary widely as a function of the design factors, ranging from 0 to almost 100%.

In order to provide a formal analysis of Type II error rates, we conducted a series of logit analyses. We used a hierarchical modeling procedure and began with a main-effects model, then introduced two-way interactions, followed by increasingly higher-order interactions. The pseudo- R^2 values for the main-effects model, based on the percentage improvement in the log-likelihood ratio, were 53%, 44%, and 31% for γ_{11} , γ_{12} , and γ_{14} , respectively (there is no Type II error for γ_{13} because the true value is zero). Although some of the higher-order interactions were significant because of the large sample size, they were substantively unimportant, as shown by the minor improvements in the pseudo- R^2 values (about 1%). Consistent with the findings for the accuracy of the coefficients and standard errors, Type II errors go

up as the level of multicollinearity increases, measure reliability deteriorates, R^2 decreases, and sample size gets smaller. For γ_{11} and γ_{12} the most important determinant of Type II error rates was multicollinearity, followed by R^2 , measure reliability (for γ_{11}) or sample size (for γ_{12}), and coefficient weight. Although coefficient weight was not a significant influence on the inaccuracy of parameter estimates, the effect on Type II error rates makes sense because the magnitude of γ_{12} is twice as high in the equal-weight condition than in the unequal-weight condition. As one would expect, Type II error rates were higher in the equal-weight condition for γ_{11} and lower in the equal-weight condition for γ_{12} . For γ_{14} , R^2 was the most important determinant of Type II error rates, followed by measure reliability, sample size, multicollinearity, and coefficient weight. Multicollinearity had a relatively small influence because ξ_4 was specified to be uncorrelated with the other constructs.

It should be emphasized that the impact of multicollinearity on Type II error rates for γ_{11} and γ_{12} was substantial. Specifically, for γ_{11} the average error rates were 8%, 17%, 39%, and 89% for increasing levels of multicollinearity, and for γ_{12} , the corresponding figures were 18%, 31%, 53%, and 91%.

Table 2 Percentage of Type II Error Rates: Experiment 1*

	Equal weights												Unequal weights											
	γ_{11}				γ_{12}				γ_{14}				γ_{11}				γ_{12}				γ_{14}			
	Collinearity level				Collinearity level				Collinearity level				Collinearity level				Collinearity level				Collinearity level			
	0.5	0.65	0.8	0.95	0.5	0.65	0.8	0.95	0.5	0.65	0.8	0.95	0.5	0.65	0.8	0.95	0.5	0.65	0.8	0.95	0.5	0.65	0.8	0.95
Sample size 3:1																								
<i>Model $R^2 = 0.25$</i>																								
REL=0.7	59.0	79.0	93.5	99.0	50.5	72.5	92.0	99.5	77.5	80.0	82.5	88.5	37.5	62.5	89.0	99.0	79.0	88.5	94.5	99.5	75.5	78.0	81.5	87.5
REL=0.8	39.5	60.0	82.0	99.5	35.5	51.5	79.5	99.0	65.0	67.5	72.0	82.0	19.0	39.5	73.5	98.5	62.5	79.5	92.0	99.0	63.5	65.0	70.0	80.5
REL=0.9	21.0	41.5	69.0	97.5	17.0	39.5	63.0	96.0	53.5	56.0	59.0	67.5	7.5	17.5	48.5	95.5	47.5	65.0	81.5	97.0	51.0	53.5	55.5	64.5
<i>Model $R^2 = 0.5$</i>																								
REL=0.7	26.5	49.0	82.0	98.5	20.5	42.5	77.0	99.0	47.0	52.0	62.5	76.0	8.5	28.5	68.0	98.5	50.0	65.0	88.5	99.0	44.5	49.5	58.5	74.5
REL=0.8	8.0	24.5	62.0	98.0	4.5	20.0	55.0	96.0	23.0	26.0	33.0	57.0	1.0	8.0	39.0	96.0	28.5	50.0	74.0	96.5	21.0	25.0	30.0	54.5
REL=0.9	1.5	4.5	29.0	91.0	0.0	4.5	28.0	90.5	12.0	14.0	17.0	27.5	0.0	0.5	11.0	87.0	11.0	28.5	56.0	93.5	9.0	10.5	12.5	25.0
<i>Model $R^2 = 0.75$</i>																								
REL=0.7	9.0	24.0	60.5	95.5	6.0	22.5	57.0	95.5	16.0	20.5	28.0	54.5	0.5	11.5	50.0	94.5	28.5	50.0	73.5	98.0	16.5	18.5	25.5	52.0
REL=0.8	0.5	6.5	30.5	92.5	0.0	4.0	29.0	89.5	5.0	5.5	6.5	27.5	0.0	0.5	16.0	90.5	8.0	23.5	56.0	91.0	4.5	5.0	7.0	26.5
REL=0.9	0.0	0.5	3.5	78.5	0.0	0.0	4.0	74.0	0.5	0.5	1.0	4.5	0.0	0.0	1.0	66.5	0.5	4.5	24.5	84.0	0.0	0.0	0.5	5.5
Sample size 6:1																								
<i>Model $R^2 = 0.25$</i>																								
REL=0.7	26.0	52.0	76.0	99.5	27.0	48.0	73.0	98.0	49.0	52.0	57.5	76.5	7.5	27.0	66.0	99.0	52.5	68.0	88.0	98.0	48.0	50.0	54.0	74.5
REL=0.8	9.5	29.5	64.5	96.5	11.0	28.0	57.5	97.0	33.5	36.0	39.0	61.0	2.0	10.5	47.0	94.5	42.5	56.8	77.5	98.5	32.0	35.0	36.5	55.5
REL=0.9	3.0	13.0	42.5	87.5	2.5	15.5	42.5	90.0	21.0	22.2	23.0	28.0	1.0	2.5	21.5	84.0	24.0	42.5	64.5	93.5	18.5	19.5	22.5	25.5
<i>Model $R^2 = 0.5$</i>																								
REL=0.7	2.5	15.0	53.5	97.0	3.0	19.5	52.5	95.5	13.5	16.0	22.0	51.0	0.5	4.0	36.0	96.5	28.0	46.5	72.0	96.0	12.5	15.0	19.0	47.5
REL=0.8	1.0	3.0	29.0	91.0	0.0	3.5	31.5	89.0	5.5	7.0	8.5	20.5	0.0	0.5	9.5	90.0	7.5	28.5	56.0	93.5	4.0	5.5	8.0	19.5
REL=0.9	0.0	1.0	6.5	77.5	0.0	0.0	9.0	73.0	0.0	0.5	1.5	2.5	0.0	0.0	1.5	66.5	0.5	9.0	33.0	82.0	0.0	0.0	0.0	0.5
<i>Model $R^2 = 0.75$</i>																								
REL=0.7	0.0	3.0	34.0	94.0	0.0	3.0	34.5	90.5	3.0	3.0	3.5	28.5	0.0	0.0	12.5	91.5	6.5	28.5	59.5	93.0	2.5	2.5	4.0	23.5
REL=0.8	0.0	0.0	7.0	86.5	0.0	0.0	7.0	78.5	0.0	0.0	0.0	5.0	0.0	0.0	1.0	80.0	0.0	4.5	33.0	86.0	0.0	0.0	0.0	6.0
REL=0.9	0.0	0.0	0.0	54.0	0.0	0.0	0.0	56.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	37.0	0.0	0.0	5.5	72.0	0.0	0.0	0.0	0.0

Note. REL, Reliability.

* We use the following shading protocol: Type II error rate of less than 19.5% no shading and then the darkness increases in the range 20%–39.5%, 40%–59.5%, 60%–79.5%, and greater than 80%.

4.2. Experiment 2

We had to generate 14,824 samples (an additional 3%) in order to obtain 200 proper solutions in each cell of the design. Most of the 424 improper solutions (86%) occurred in the lowest reliability condition. Improper solutions were particularly likely when reliability was low, sample size was small, and R^2 was high. Thus, while the first study showed that models with very high levels of multicollinearity are prone to improper solutions, the second study indicates that poor measure reliability is another major determinant of improper solutions. The difference in results between the two studies can be attributed to the different levels of multicollinearity and reliability used in the simulations (the highest level of multicollinearity was 0.95 in Experiment 1 and 0.8 in Experiment 2; the lowest level of reliability was 0.7 in Experiment 1 and 0.6 in Experiment 2).

The design factors explained 20%, 21%, 17%, 16%, 17%, 12%, 13%, and 9% of the variance in the accuracy of the coefficients and 83%, 84%, 86%, 71%, 74%, 81%, 83%, and 84% of the variance in the accuracy of the standard errors for γ_{11} , γ_{12} , γ_{13} , γ_{21} , γ_{22} , γ_{23} , β_{31} , and β_{32} , respectively (Table 1). Again, the main effects of

multicollinearity, measurement error, R^2 , and sample size accounted for most of the explained variance in the coefficients and standard errors (around 96%). In particular, multicollinearity, along with measure unreliability, was among the two most important determinants of inaccurate coefficients and standard errors for γ_{11} and γ_{12} , which is as expected because multicollinearity was manipulated by varying the correlation between ξ_1 and ξ_2 . In contrast, multicollinearity exerted relatively minor effects on γ_{13} , γ_{23} , β_{31} , and β_{32} because ξ_3 was specified to be uncorrelated with the other exogenous constructs, and the correlation between η_1 and η_2 was generally small. Thus, when there are multiple endogenous variables, it seems that the damaging effects of multicollinearity at one stage of the model (between ξ_1 and ξ_2) do not transfer to another stage, as long as the correlation between constructs at the next stage is small.

The design factors again resulted in a significant percentage of Type II errors, which can be as high as 94% (Table 3). Based on logit analyses, the main-effects model yielded pseudo- R^2 values of 42%, 39%, 38%, 35%, 42%, and 33% for γ_{11} , γ_{12} , γ_{21} , γ_{23} , β_{31} , and β_{32} , respectively. Higher-order interactions, although

Table 3 Percentage of Type II Error Rates: Experiment 2*

		γ_{11}			γ_{12}			γ_{21}			γ_{23}			β_{31}			β_{32}		
		Collinearity level			Collinearity level			Collinearity level			Collinearity level			Collinearity level			Collinearity level		
		0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8
Sample size 3:1																			
$R^2 = 0.25$	REL=0.6	41.0	68.5	90.0	62.0	78.0	91.0	52.5	71.0	93.5	21.5	24.5	39.5	32.5	29.0	27.5	33.5	40.5	43.5
	REL=0.7	20.5	44.0	85.5	39.5	65.0	83.5	26.5	46.0	80.5	7.5	7.5	10.0	14.5	11.5	10.5	14.0	19.5	22.5
	REL=0.8	8.0	25.0	65.5	22.0	45.5	76.0	12.0	27.5	58.0	2.0	2.0	2.5	3.0	2.5	2.0	6.5	8.0	10.0
	REL=0.9	2.5	12.0	45.5	9.5	27.0	61.0	1.5	7.0	38.0	0.5	0.5	0.5	0.0	0.0	0.0	2.0	2.0	3.0
$R^2 = 0.50$	REL=0.6	12.0	34.5	82.0	25.0	53.5	83.5	35.5	53.5	83.5	7.0	6.0	15.5	6.5	5.5	5.0	9.0	12.0	15.0
	REL=0.7	0.5	13.0	58.5	6.5	30.5	66.5	9.0	18.0	60.0	0.0	0.0	0.5	0.0	0.0	0.0	3.0	3.0	5.5
	REL=0.8	0.0	1.5	32.5	0.5	8.0	50.5	1.0	2.0	28.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.5
	REL=0.9	0.0	0.0	10.5	0.0	0.0	26.5	0.0	0.5	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$R^2 = 0.75$	REL=0.6	2.5	20.5	63.0	9.0	26.5	69.0	56.0	60.5	87.0	3.5	4.0	8.5	0.5	0.5	2.0	5.0	5.0	10.0
	REL=0.7	0.0	1.5	37.0	0.0	7.0	46.0	25.0	28.0	54.0	1.0	2.0	2.0	0.0	0.0	0.0	1.5	2.5	3.5
	REL=0.8	0.0	0.0	9.5	0.0	0.5	20.0	2.5	3.5	13.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	REL=0.9	0.0	0.0	0.5	0.0	0.0	1.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sample size 6:1																			
$R^2 = 0.25$	REL=0.6	11.0	35.5	78.0	22.5	53.0	84.0	13.5	30.0	73.0	0.5	0.5	5.5	1.5	1.5	1.0	5.5	7.5	9.0
	REL=0.7	2.5	18.0	61.0	7.5	31.5	72.5	3.0	10.5	52.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.5	1.0
	REL=0.8	1.0	3.5	42.5	1.5	13.0	56.0	0.0	2.5	28.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	REL=0.9	0.0	1.0	19.5	0.5	3.5	33.0	0.0	0.0	9.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$R^2 = 0.50$	REL=0.6	1.5	9.0	56.0	4.0	17.0	61.5	4.5	9.5	57.5	0.0	0.0	1.5	0.0	0.0	0.0	0.5	0.5	0.5
	REL=0.7	0.0	1.5	32.0	0.0	5.0	42.0	0.0	1.0	20.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	REL=0.8	0.0	0.5	10.0	0.0	0.5	18.5	0.0	0.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	REL=0.9	0.0	0.0	0.5	0.0	0.0	3.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$R^2 = 0.75$	REL=0.6	0.0	3.0	38.0	0.0	5.5	43.5	19.5	23.0	52.5	0.0	0.5	1.0	0.0	0.0	0.0	0.0	0.5	0.5
	REL=0.7	0.0	0.5	13.0	0.0	0.0	18.0	4.5	2.5	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	REL=0.8	0.0	0.0	0.5	0.0	0.0	3.5	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	REL=0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Note. REL, Reliability.

* We use the following shading protocol: Type II error rate of less than 19.5% no shading and then the darkness increases in the range 20%–39.5%, 40%–59.5%, 60%–79.5%, and greater than 80%.

significant because of the large sample, did not explain substantively important portions of variance (less than 2%). Due to the limited number of Type II errors for γ_{23} , β_{31} , and β_{23} in certain cells, higher-order interactions (and the main-effects model for β_{31}) could not be estimated reliably, but visual inspection of the data indicated that there was little evidence of interaction effects. In general, Type II errors increase as the level of multicollinearity increases, measure reliability deteriorates, R^2 decreases, and sample size goes down. For γ_{11} and γ_{12} (which are most affected by high correlation between ξ_1 and ξ_2), multicollinearity was the major determinant of Type II error rates, followed by R^2 , measure reliability, and sample size. Multicollinearity was a less important influence on Type II error rates for γ_{21} , and it had little effect in the case of γ_{23} , β_{31} , and β_{32} . Measure reliability was the major determinant of Type II error rates for γ_{21} , γ_{23} , and β_{32} . The Type II error rates for γ_{11} and γ_{12} as a function of multicollinearity were nontrivial, although not as high as in Experiment 1 since the highest level of multicollinearity was only 0.8 (rather than 0.95).

5. General Discussion

The simulations show that high multicollinearity, in combination with low measure reliability, small sample size, and low explained variance in endogenous constructs, has several undesirable consequences. To begin with, the researcher may not be able to get a proper solution. Particularly in models with very high levels of collinearity (correlations among the exogenous variables greater than 0.9) and low measure reliability (composite reliability smaller than 0.7), improper solutions were quite common. Thus, improper solutions might be a signal of multicollinearity (and other model problems) and researchers should pay careful attention to them. Even when a proper solution can be obtained, multicollinearity can lead to inaccurate parameter estimates and a high incidence of Type II errors, particularly when reliability is weak, sample size is small, and explained variance is low.

Two things should be kept in mind, however. First, the adverse effects of multicollinearity on estimation accuracy occurred primarily at the most extreme levels of multicollinearity between ξ_1 and ξ_2 . Consistent with the findings of Mason and Perreault (1991), in the first experiment the highest correlation between ξ_1 and ξ_2 was 0.95, and it was in this condition that the inaccuracy in the estimates of the path coefficients and standard errors was most pronounced. In the second experiment, in which the highest correlation between ξ_1 and ξ_2 was “only” 0.80, multicollinearity, although still an important influence on the accuracy of the estimation results, generally played a less pronounced role (Table 1). Even in this case, most of

the adverse effects of multicollinearity occurred at the highest level of multicollinearity. It should be noted, however, that high levels of multicollinearity may be less apparent in SEM than regression. Because of the attenuation created by measurement error, the cross-construct correlations among indicators will be lower than the actual level of multicollinearity (i.e., correlations among the exogenous constructs). For example, in the first experiment the correlation between ξ_1 and ξ_2 was 0.95, but the correlation between the indicators of ξ_1 and ξ_2 was only 0.35 when reliability was 0.7, 0.48 when reliability was 0.8, and 0.66 when reliability was 0.9. Thus, correlations between the observed variables that look innocuous may induce fairly high levels of multicollinearity among the latent constructs.

Second, as pointed out by Mason and Perreault (1991), multicollinearity should not be viewed in isolation. It is also important to consider other factors that influence the accuracy of estimation results, and thus may either exacerbate or mitigate the harmful effects of multicollinearity. Consistent with Mason and Perreault (1991), our results showed that the deleterious effects of multicollinearity could be largely offset when the sample size was large and the independent variables explained a high proportion of the variance in the dependent variable. Extending the results of Mason and Perreault (1991), our findings show that measure reliability is another important influence on estimation accuracy. In fact, measure reliability was generally either the most important or second-most important determinant of the accuracy of the estimated coefficients and standard errors of γ_{11} and γ_{12} . Thus, researchers should not assume that because SEM takes measurement error into account and corrects paths for attenuation, measure unreliability is less of a problem. The present findings clearly show that whenever there is unreliability of measurement, estimation accuracy for both coefficients and standard errors suffers. In contrast, when measures are highly reliable, even fairly high levels of multicollinearity can be tolerated. An added benefit is that when reliability is high, correction for attenuation will not increase the correlation between latent constructs to unacceptably high levels.

The major consequence of inaccurate estimates of coefficients and standard errors is that the likelihood that a researcher will commit a Type II error can be substantial. The Type II error rates tabulated in Tables 2 and 3 for the various conditions of both experiments suggest the following conclusions:

- When multicollinearity is extreme (around 0.95), Type II error rates are generally unacceptably high (over 80%).
- When multicollinearity is between 0.6 and 0.8, Type II error rates can be substantial (greater than 50% and frequently above 80%) when composite reliability

is weak (0.7 or lower), R^2 is low (0.25), and sample size is relatively small (ratio of 3:1). However, as reliability improves (0.80 or higher), R^2 reaches 0.75, and sample becomes relatively large (ratio of 6:1), Type II error rates become negligible (below 5%).

- When multicollinearity is between 0.4 and 0.5, Type II error rates tend to be quite small, except when reliability is weak (0.7 or below), R^2 is low (0.25), and sample size is small (ratio of 3:1), in which case error rates can still be high (greater than 50%).

One interesting finding that emerged in Experiment 1 was that the Type II error rates in the 0.9 reliability conditions were sometimes substantially higher than those reported by Mason and Perreault (1991), whose simulations assumed perfect reliability (which is highly unrealistic in practice). As an example, consider the cell in which multicollinearity is 0.8 to 0.95, R^2 is 0.75, the weights of γ_{11} and γ_{12} are equal, and the sample size is relatively small (150 observations). In this case, Mason and Perreault (1991) reported Type II error rates of 14% and 15% for γ_{11} and γ_{12} , respectively. We replicated these findings using our data-generation mechanism and SEM software (our figures were 15% and 16%, respectively). However, when reliability decreased to 0.9 (which most researchers would probably consider excellent), the error rates were 79% and 74%, respectively (Table 2). This vividly demonstrates the importance of reliable measurement for valid tests of theories, particularly when multicollinearity is high.

5.1. Avoiding Problems with Estimates Due to Multicollinearity

Because remedies for multicollinearity after the fact are of limited value, it would be wise for researchers to inoculate their studies against the harmful effects of excessive shared variance among the predictors. One recommendation is to include all-important influences on a dependent variable so that the R^2 's for all endogenous constructs are relatively high. Unfortunately, behavioral models often have low explanatory power (Peterson et al. 1985), so that this may not be an option. A more viable suggestion is to make sure that the sample size is adequate. In planning a study, researchers are admonished to consider the issue of expected effect sizes and determine the sample size needed to achieve reasonable levels of power (Cohen 1992, MacCallum et al. 1996).

Probably the most important safeguard against the damaging effects of multicollinearity is to make sure that all constructs are measured as reliably as possible. It appears that there is no discrete cut-off between good and bad reliability. As reliability increases (even from 0.8 to 0.9 or higher), estimation accuracy improves and Type II errors go down. This finding means that researchers have to devote a lot

of attention to measure development (Rossiter 2003). Prior to conducting a study, the construct has to be defined clearly, and then appropriate items have to be sampled from its domain. Because three to five good indicators of each construct should be available, pre-studies are necessary to make sure that all items load highly on the intended construct, or additional measures should be included in the study so that a sufficient number of reliable items can be retained. Special care is necessary to distinguish reflective and formative measurement models, and to specify only reflective indicators as functions of underlying factors, because formative indicators are unlikely to achieve high internal consistency (Jarvis et al. 2003).

If these guidelines are not followed, there can be serious implications for theory testing, and even modest levels of multicollinearity will lead to high Type II errors rates. When Type II errors occur, researchers can easily derive misleading conclusions from their data (e.g., eliminate important variables from a model, ascribe the wrong sign or effect size to a variable), and inconsistencies across studies are likely.

5.2. Diagnosing and Controlling Potential Multicollinearity Problems

Commonly used methods for assessing the degree of multicollinearity, such as bivariate correlations, variance inflation factors, and condition numbers, were briefly discussed earlier.⁷ Here we will focus on a diagnostic that has not been discussed in the literature. The problem of multicollinearity is closely related to the issue of discriminant validity. If constructs are too highly correlated, they lack discriminant validity. Researchers who use SEM usually conduct measurement analyses prior to testing structural relationships, and often assess discriminant validity by testing whether the correlations (corrected for measurement error) among constructs differ from one. If this is not the case, multicollinearity is probably extreme, and the researcher will most

⁷ For both the experiments, we examined the efficacy of condition numbers in signaling the cases impacted by multicollinearity. In Experiment 1, the mean condition number across all samples was 4.10 (median of 2.95). The maximum condition number was 53.16. Condition numbers greater than 30 occurred only in the highest multicollinearity condition. The correlations between the condition numbers and the Type II error rates for γ_{11} , γ_{12} , and γ_{14} were 0.55, 0.48, and 0.22. In Experiment 2, the mean condition number across all samples was 2.25 (median of 2.02). The maximum condition number was 10.94. The correlations between the condition numbers and the Type II error rates for γ_{11} and γ_{12} were 0.46 and 0.44. Belsley et al. (1980) suggest that condition indices of 5–10 indicate weak dependencies, and indices greater than 30 indicate moderate to strong dependencies. Thus, based on the condition numbers there are no strong dependencies in Experiment 2, but there are strong dependencies in the highest multicollinearity condition in Experiment 1.

likely respecify the model because the distinct conceptual status of the constructs in question is questionable (Anderson and Narus 1984). It is therefore unlikely that structural equation models will contain constructs that are very highly correlated. However, correlations in the 0.7 or 0.8 range are fairly common, and they will probably be distinct from one. Fornell and Larcker (1981a) have proposed another test of discriminant validity that is more demanding and involves comparing the squared correlation between two constructs to the average variance extracted statistic. The idea is that a construct should be more closely related to its own indicators than to other constructs. Since average variance extracted is a measure of reliability, and since multicollinearity and reliability are the two major influences on estimation accuracy and inference errors, the question arises whether the Fornell and Larcker discriminant validity criterion can be used to flag cases of severe multicollinearity. Using the data from the second experiment, we calculated whether ξ_1 and ξ_2 achieved discriminant validity and cross-classified this variable with Type II errors for γ_{11} and γ_{12} . Overall, the correlations between the two variables for γ_{11} and γ_{12} were -0.43 and -0.38 , respectively. Thus, as expected inference errors were less likely when discriminant validity was achieved. Specifically, when the two constructs were distinct, the Type II error rates were only 5% and 9% for γ_{11} and γ_{12} , respectively. On the other hand, when ξ_1 and ξ_2 lacked discriminant validity, a Type II error occurred in 39% and 42% of the cases, respectively. These findings show that if the Fornell and Larcker criterion is satisfied, an inference error is unlikely. However, when the criterion is not satisfied, it does not necessarily mean that the chance of committing a Type II error is high.

If multicollinearity does create estimation problems, we recommend that SEM be used to conduct a measurement analysis and to obtain a covariance matrix among the latent constructs. The estimated covariance matrix can then be used as input into other data-analytic techniques that are more robust to multicollinearity, such as ridged partial least squares (Jagpal 1982). Preliminary evidence suggests that estimates of bivariate relationships between constructs are not adversely affected by multicollinearity. Specifically, we conducted an analysis in which CFA models were fit to the data from Experiment 2. We then compared the estimated correlations (as indicated in the standardized Φ matrix) with the population values. Out of 15 possible comparisons, multicollinearity was significant in only five cases, and even then it accounted for at most 3% of the inaccuracy in the estimated correlations. Thus, as one might have expected a priori, multicollinearity does not seem to adversely affect the accuracy of estimated *bivariate* relationships among constructs.

5.3. Limitations

Because of the complexity of structural equation models, there are many possible factors that might affect parameter estimation under conditions of multicollinearity. Our goal was to examine some basic factors affecting simple models. However, to fully understand the degree to which multicollinearity will affect estimates, additional factors and levels within these factors would need to be examined. For example, we have only examined models in which all indicators have similar levels of reliability. It is unclear how differences in measure reliability across constructs affect estimation, or whether nonuniform loadings on a construct lead to different results. Our findings will also have to be extended to other types of models. Complexities such as multigroup models, nonnormal distributions, nonrecursive models, and second-order factors may further complicate the effects of multicollinearity. In order to fully document the effects of multicollinearity, it may be necessary to develop a host of tables that are unique to various model conditions and structures.

5.4. Conclusion

We attempted to identify conditions that allow researchers to avoid the problems associated with multicollinearity in SEM. Our main conclusion is that good measure reliability, a model whose explanatory power is high, and a large sample size can effectively protect against the deleterious effects of multicollinearity unless the multicollinearity is severe. Since multicollinearity is not easily remedied once it is present (it is ultimately a data problem, in the sense that two or more “independent” variables do not have sufficient independent variation, rather than an estimation problem), and since deficiencies on other factors lead to nonnegligible inference errors even when multicollinearity is not particularly high, it behooves the researcher to pay careful attention to the conditions known to mitigate multicollinearity problems prior to conducting a study.

In the past, researchers have often assumed that because SEM takes into account measurement error and corrects paths for attenuation, measure unreliability is less of a problem. Our findings clearly show that this assumption is not warranted. Although SEM is better than techniques such as regression, which do not consider measurement error and therefore lead to inconsistent parameter estimates, measurement error strongly affects estimation accuracy for both coefficients and standard errors, and, as a consequence, increases the likelihood of Type II errors. Even when reliability is fairly high by conventional standards, measurement error can be damaging. Thus, nothing substitutes for good quality measures, and researchers should make every attempt to use reliable measures

that are discriminant from other constructs. Of course, the worst scenario is a situation in which a single item is employed to measure a construct and the presumed objectivity of the measure (e.g., the amount of sales and expenditures or other kinds of secondary data) is used to justify the lack of multi-item measurement. In this case, unreliability of measurement cannot even be assessed, and the damaging effects of unreliability are completely hidden.

Acknowledgments

The authors contributed equally. The authors appreciate helpful feedback from Donald R. Lehmann (Columbia University) and Edward E. Rigdon (Georgia State University). The article benefited from the feedback of the editor, the associate editor, and three anonymous *Marketing Science* reviewers.

References

- Anderson, James C., David W. Gerbing. 1984. The effects of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika* **49** 155–173.
- Anderson, James C., James A. Narus. 1984. A model of the distributor's perspective of distributor-manufacturer working relationships. *J. Marketing* **48**(Fall) 62–74.
- Ansari, Asim, Kamel Jedidi, Sharan Jagpal. 2000. A hierarchical Bayesian methodology for treating heterogeneity in structural equation models. *Marketing Sci.* **19**(4) 328–347.
- Bagozzi, Richard P., Hans Baumgartner. 1994. The evaluation of structural equation models and hypothesis testing. *Principles of Marketing Research*. Blackwell Publishers, Oxford, England, 386–422.
- Baumgartner, Hans, Christian Homburg. 1996. Applications of structural equation modeling in marketing and consumer research: A review. *Internat. J. Res. Marketing* **13** 139–161.
- Belsley, D., E. Kuh, R. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, New York.
- Bentler, Peter M., Eric J. C. Wu. 1995. *EQS for Windows*. Multivariate Software, Inc., Encino, CA.
- Bollen, Kenneth A. 1989. *Structural Equations and Latent Variables*. John Wiley & Sons Inc., New York.
- Cohen, Jacob. 1992. A power primer. *Psych. Bull.* **112**(July) 155–159.
- Cote, Joseph A., M. Ronald Buckley. 1987. Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *J. Marketing Res.* **24**(August) 315–318.
- Cote, Joseph A., M. Ronald Buckley. 1988. Measurement error and theory testing in consumer research: An illustration of the importance of construct validation. *J. Consumer Res.* **14**(March) 579–582.
- Farley, John U., Donald R. Lehmann, Lane H. Mann. 1998. Designing the next best study for maximum impact. *J. Marketing Res.* **35**(November) 496–501.
- Fornell, Claes, David F. Larcker. 1981a. Evaluating structural equation models with unobservable variables and measurement error. *J. Marketing Res.* **28**(February) 39–50.
- Fornell, Claes, David F. Larcker. 1981b. Structural equation models with unobservable variables and measurement error: Algebra and statistics. *J. Marketing Res.* **28**(August) 382–388.
- Grapentine, Terry. 2000. Path analysis vs. structural equation modeling. *Marketing Res.* **12**(3) 12–20.
- Heise, D. R. 1969. Problems in path analysis and causal inference. E. F. Borgatta, ed. *Sociological Methodology*. Jossey-Bass, San Francisco, CA.
- Hulland, John, Yiu Ho Chow, Shunyin Lam. 1996. Use of causal models in marketing research: A review. *Internat. J. Res. Marketing* **13**(April) 181–197.
- Jagpal, Harsharanheet S. 1982. Multicollinearity in structural equation models with unobservable variables. *J. Marketing Res.* **19**(November) 431–439.
- Jarvis, Cheryl Burke, Scott B. MacKenzie, Philip M. Podsakoff. 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *J. Consumer Res.* **30**(June) 199–218.
- Jedidi, Kamel, Harsharanjeet S. Jagpal, Wayne S. DeSarbo. 1997. Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Sci.* **16**(1) 39–59.
- Kaplan, David. 1994. Estimator conditioning diagnostics for covariance structure models. *Sociological Methods Res.* **23**(November) 200–229.
- Kennedy, Peter. 1992. *A Guide to Econometrics*. MIT Press, Cambridge, MA.
- Kenny, David A. 1979. *Correlation and Causality*. Wiley, New York.
- Krishnamurthi, Lakshman, Arvind Rangaswamy. 1987. The equity estimator for marketing research. *Marketing Sci.* **6**(Fall) 336–357.
- MacCallum, R. C., M. W. Browne, H. M. Sugawara. 1996. Power analysis and determination of sample size for covariance structure modeling. *Psych. Methods* **1**(2) 130–149.
- Malhotra, Naresh K., Mark Peterson, Susan Bardi Kleiser. 1999. Marketing research: A state-of-the-art review and directions for the twenty-first century. *J. Acad. Marketing Sci.* **27**(2) 160–183.
- Maruyama, Geoffrey M. 1998. *Basics of Structural Equation Modeling*. Sage, Thousand Oaks, CA.
- Mason, Charlotte H., William D. Perreault Jr. 1991. Collinearity, power, and interpretation of multiple regression analysis. *J. Marketing Res.* **28**(August) 268–280.
- Paxton, Pamela, Patrick J. Curran, Kenneth A. Bollen, Jim Kirby, Feinian Chen. 2001. Monte Carlo experiments: Design and implementation. *Structural Equation Model.* **8** 287–312.
- Peterson, Robert A., Gerald Alba, Richard F. Beltramini. 1985. A meta-analysis of effect sizes in consumer behavior experiments. *J. Consumer Res.* **12**(June) 97–103.
- Rindskopf, David. 1984. Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods Res.* **13**(August) 109–119.
- Rossiter, John R. 2003. The C-OAR-SE procedure for scale development in marketing. *Internat. J. Res. Marketing* **19**(4) 305–335.
- Schmidt, Peter, Edward N. Muller. 1978. The problem of multicollinearity in a multistage causal alienation model: A comparison of ordinary least squares, maximum-likelihood, and ridge estimators. *Quality Quantity* **12** 267–297.
- Steenkamp, Jan-Benedict E. M., Hans Baumgartner. 2000. On the use of structural equation models for marketing modeling. *Internat. J. Res. Marketing* **17** 195–202.
- Verbeke, Willem, Richard P. Bagozzi. 2000. Sales call anxiety: Exploring what it means when fear rules a sales encounter. *J. Marketing* **64**(July) 88–101.