

On Asymptotic Robustness of NT Methods with Missing Data

Victoria Savalei  
University of California, Los Angeles

Abstract

Literature on asymptotic robustness of normal theory (NT) methods outlines conditions under which the NT estimator remains asymptotically efficient and the NT test statistic retains its chi-square distribution even under nonnormality. These conditions have been stated both abstractly and in terms of properties of specific models. This research discusses issues associated with extending asymptotic robustness theory to the direct ML estimator and associated test statistic when data are missing completely at random (MCAR). It is shown that the same abstract robustness condition necessary for robustness to hold with complete data is required for incomplete data, while properties of specific models (such as mutual independence of the errors and their independence of the factors in a CFA model) no longer ensure robustness with incomplete data. The lack of robustness in such a case is illustrated both mathematically and empirically via a simulation study. Violation becomes more severe when the data are highly nonnormal and when a higher proportion of data is missing.

## On Asymptotic Robustness of NT Methods with Missing Data

In structural equation modeling (SEM), assumptions are often made about the distribution of the data in order to obtain parameter estimates, their standard errors, and a global test of model fit. The most common assumption is that the data come from a distribution with no excess kurtosis, such as the multivariate normal. If this assumption is met, the corresponding normal theory (NT) parameter estimates are asymptotically efficient, and the associated NT test statistic is asymptotically chi-square distributed. Because of the highly-structured nature of the 4<sup>th</sup> order moments under multivariate normality, the NT test statistic converges to a chi-square variate relatively quickly, so that a chi-square approximation works well starting at medium sample sizes (e.g., Bentler & Yuan, 1999). However, more often than not, the assumption of no excess kurtosis in the population is not true, yet the above method is still utilized because of software limitations, lack of better alternatives, or other reasons. In this case, the NT parameter estimates remain consistent but are in general no longer asymptotically efficient, and therefore the NT standard errors no longer accurately reflect their variability. The NT test statistic is now asymptotically a mixture of independent one degree of freedom chi-square variates, with weights depending on both the properties of the model and on the kurtosis matrix of the population (Satorra, 1989; Satorra & Bentler, 1994).

One solution is to rescale the NT test statistic by the estimated sum of the mixture weights, leading to a statistic whose distribution is better approximated by a chi-square variate (Satorra & Bentler, 1994). Correct standard errors for NT parameter estimates can be obtained by computing the so called “sandwich” standard errors (Satorra & Bentler, 1994), which reflect the greater variability of the NT parameter estimates under nonnormality. This approach is a viable option. However, researchers have also identified conditions under

which NT parameter estimates remain efficient even under nonnormality and the NT test statistic converges to a chi-square variate without any corrections. These conditions have been stated in abstract form (e.g., Shapiro, 1986) as well as in terms of properties of specific models (see references below). It should be noted that while these conditions ensure the estimates remain asymptotically efficient, NT standard errors will in general no longer be accurate, except for some parameters (Amemiya, Fuller, & Anderson, 1987; Anderson & Amemiya, 1988). Sandwich-type standard errors will still be correct, however. Thus, asymptotic robustness of NT methods is primarily of interest because it ensures the correct limiting distribution of the NT test statistic. Asymptotic robustness of the NT methods has been studied in the context of the factor analytic models and more general LISREL models (Amemiya & Anderson, 1990; Browne, 1987; Browne & Shapiro, 1988; Mooijaart & Bentler, 1991; Shapiro, 1987), growth curve models (Browne, 1990; Satorra, 2001), multiple-group mean and covariance structure models (e.g., Satorra, 2001, 2002), and multilevel models (Yuan & Bentler, 2005). It should be noted that the robustness of other statistics can also be studied, even if those that are not chi-square distributed (Satorra & Bentler, 1990), but NT methods remain the most important practical application in asymptotic robustness literature.

Nonnormal data is only one of the complications that SEM researchers encounter. Missing data is another problem that complicates statistical inference, perhaps equally as vexing. Real life data often combine both of these characteristics: they are nonnormal as well as incomplete. NT methods have been successfully extended to incomplete data whenever the missing mechanism is such that a consistent unstructured estimate of the population covariance matrix is available (Yuan & Bentler, 2000). However, their asymptotic robustness properties have not been studied. The goal of this paper is to investigate the robustness properties of the NT estimators and the NT test statistic when the data are incomplete as well

as nonnormal. We focus on covariance structure models and use the factor analytic model as our main example.

This paper is organized as follows. In Section 1, the necessary asymptotic theory and abstract conditions ensuring robustness of NT methods with complete data are reviewed. Instead of reviewing the mathematics behind the many existing approaches (e.g., Amemiya & Anderson, 1990; Browne & Shapiro, 1988; Mooijaart & Bentler, 1991), we adopt the framework of Mooijaart and Bentler (1991), which will also be used to study what happens with incomplete data. In Section 2, the necessary asymptotic theory for incomplete data is reviewed and some additional results are derived. It is then shown that the abstract condition required for robustness of NT methods to hold with complete data generalizes fully to incomplete data. In Section 3, the connection between complete and incomplete data is explored. The following question is asked: if incomplete data were obtained from complete data where robustness holds, will robustness hold in the resulting incomplete data population? It is shown that the answer is, unfortunately, no. This result implies that conditions on specific models that ensure asymptotic robustness with complete data will not generalize to incomplete data. This is illustrated in Section 4, where specific conditions for robustness under a general factor model are stated, and an example is given to show that these conditions do not ensure robustness with incomplete data. The resulting lack of robustness is illustrated empirically via a simulation study comparing the performance of the NT test statistic with complete and incomplete nonnormal data. Finally, Section 5 contains a discussion. We address why the multiple group approach to robustness developed by Satorra (e.g., 2001) is not the correct framework to approach the study of robustness with missing data, despite the often drawn analogy between multiple group and incomplete data situation. Implications of our results for practitioners are also discussed.

### 1. Asymptotic Robustness with Complete Data

Let the rows of the  $n \times p$  data matrix  $X$  be a random sample of size  $n$  from a population with the mean vector  $\mu(\theta)$  and a  $p \times p$  covariance matrix  $\Sigma(\theta)$ , where  $\theta$  is a  $q \times 1$  vector of parameters. That is,  $\theta$  includes both mean and covariance structure parameters. In this paper, only covariance structure models are studied, so that  $\theta$  can be partitioned as  $\theta = (\theta_1', \mu')'$ , where  $\Sigma = \Sigma(\theta_1)$ . Define a  $(p^* + p) \times 1$  vector  $\beta = (\sigma', \mu')'$ , where  $\sigma = \text{vech}\Sigma$ , the vector of nonredundant elements of  $\Sigma$  and  $p^* = .5p(p+1)$ . Let  $\bar{x} = \frac{1}{n} X'1$  be the vector of sample means, and  $S = \frac{1}{n} (X - 1\bar{x}')'(X - 1\bar{x}')$  be the sample covariance matrix. We obtain  $\hat{\beta}$ , the normal theory maximum likelihood (ML) estimate under the saturated model  $\Sigma(\beta)$ , by maximizing:

$$l(\beta) = \sum_{i=1}^n l_i(\beta) = \frac{n}{2} (p \log(2\pi) - \log |\Sigma(\beta)| - \text{tr}(S\Sigma^{-1}(\beta)) - (\bar{x} - \mu(\beta))' \Sigma^{-1}(\beta) (\bar{x} - \mu(\beta))).$$

We obtain  $\hat{\theta}$ , the normal theory ML estimate under the structured model  $\Sigma(\theta)$ , by maximizing:

$$l(\theta) = \sum_{i=1}^n l_i(\theta) = \frac{n}{2} (p \log(2\pi) - \log |\Sigma(\theta)| - \text{tr}(S\Sigma^{-1}(\theta)) - (\bar{x} - \mu(\theta))' \Sigma^{-1}(\theta) (\bar{x} - \mu(\theta))).$$

Let  $s = \text{vech}S$  be the vector of nonredundant elements of  $S$ . With complete data, the estimator  $\hat{\beta}$  has a simple form:  $\hat{\beta} = (s', \bar{x}')'$ . The likelihood ratio test statistic for testing the hypothesis  $\Sigma = \Sigma(\theta)$  against the hypothesis  $\Sigma = \Sigma(\beta)$  is given by  $T_{ML} = -2(l(\hat{\theta}) - l(\hat{\beta}))$  or equivalently  $T_{ML} = n\hat{F}_{ML} = nF_{ML}(\hat{\beta}, \beta(\hat{\theta}))$ , where  $F_{ML}(\hat{\beta}, \beta(\hat{\theta}))$  is the minimum of  $F_{ML}(\hat{\beta}, \beta(\theta)) = \text{tr}(S\Sigma^{-1}(\theta)) - \log |S\Sigma^{-1}(\theta)| + (\bar{x}_n - \mu(\theta))' \Sigma^{-1}(\theta) (\bar{x}_n - \mu(\theta)) - p$ . Thus,  $\hat{\theta}$  can equivalently be defined as the minimizer with respect to  $\theta$  of the discrepancy function  $F_{ML}$ .

Following Browne (1984), a scalar valued function  $F(a, b)$  is called a discrepancy function if it has the following three properties:  $F(a, b) \geq 0$  for all  $a, b$ ;  $F(a, b) = 0$  if and only if  $a = b$ ; and  $F$  is twice continuously differentiable in  $a$  and  $b$ . However, the definition of  $\hat{\theta}$  and  $T_{ML}$  via the likelihood equations makes the extension to missing data more straight-forward.

In the asymptotic robustness literature, robustness of GLS (generalized least squares) discrepancy functions is usually studied (e.g., Browne, 1987; Mooijaart & Bentler, 1991). A GLS discrepancy function has the form  $F_{GLS} = (\hat{\beta} - \beta(\theta))' \hat{V} (\hat{\beta} - \beta(\theta))$  for some positive definite matrix  $\hat{V}$ , possibly random. A discrepancy function not of this form, such as  $F_{ML}$ , can be shown to asymptotically have the same minimizer, and the same minimum, as an appropriately chosen GLS discrepancy function (Browne, 1974; Shapiro, 1985). Let  $\theta_0$  be the true parameter value, and let  $\beta_0 = \beta(\theta_0)$ . Shapiro (1985) showed that under mild regularity conditions which ensure that  $\hat{\beta} \xrightarrow{p} \beta_0$ , the minimizer (and the minimum) of any discrepancy function  $F$  and the minimizer (and the minimum) of

$F_{GLS} = (\hat{\beta} - \beta(\theta))' \hat{V} (\hat{\beta} - \beta(\theta))$ , where  $\hat{V}$  is any consistent estimate of

$V = \frac{1}{2} \ddot{F}(\beta_0) = \frac{1}{2} \left. \frac{\partial^2 F(\beta)}{\partial \beta \partial \beta'} \right|_{\beta_0}$  are asymptotically equivalent. It follows that the respective test

statistics,  $n\hat{F}$  and  $n\hat{F}_{GLS}$ , are asymptotically equivalent. Thus, to study asymptotic properties of the ML estimator and test statistic, it is sufficient to study asymptotic properties of the appropriately chosen GLS estimator and test statistic. In particular,  $F_{ML}$  has the following

Hessian matrix:  $\frac{1}{2} \ddot{F}_{ML}(\beta_0) = \begin{pmatrix} \frac{1}{2} D_p' (\Sigma^{-1} \otimes \Sigma^{-1}) D_p & 0 \\ 0 & \Sigma^{-1} \end{pmatrix}$ , where  $D_p$  is the duplication

matrix<sup>1</sup> of Magnus and Neudecker (1999), and the blocks correspond to the partitioning  $\beta = (\sigma', \mu')'$ . From this result, we have the asymptotic equivalence of the ML

approach and, for example, a GLS approach that sets  $\hat{V} = \begin{pmatrix} \frac{1}{2}D_p'(S^{-1} \otimes S^{-1})D_p & 0 \\ 0 & S^{-1} \end{pmatrix}$ . All

approaches asymptotically equivalent to the ML approach will be referred to as NT.

We also need the notion of a correctly specified discrepancy function. The central limit theorem allows us to assume that  $\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, \Gamma)$  for some positive definite matrix  $\Gamma$ . The discrepancy function  $F$  is said to be *correctly specified* if  $\Gamma^{-1} = \frac{1}{2}\ddot{F}(\beta_0)$  (Browne, 1984). In other words, the discrepancy function  $F$  is correctly specified if the minimizer  $\hat{\theta}$  of  $F$  and the minimum value  $\hat{F} = F(\hat{\theta})$  are asymptotically equivalent to the minimizer and the minimum of a GLS function with the weight matrix  $\hat{V}$  such that  $\hat{V} \xrightarrow{p} \Gamma^{-1}$ .

For a correctly specified discrepancy function,  $\hat{\theta}$  is asymptotically efficient, with

$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, (\Delta'\Gamma^{-1}\Delta)^{-1})$  where  $\Delta = \left. \frac{\partial\beta(\theta)}{\partial\theta'} \right|_{\theta_0}$ , and the test statistic  $n\hat{F}$  is

asymptotically chi-square distributed with  $p^* + p - q$  degrees of freedom.

The discrepancy function can be correctly specified in two ways. Some true assumptions can be made about  $\Gamma$  and the discrepancy function can be chosen in accordance with those assumptions. For example, if the data are normal, and we correctly assume that

$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, \Gamma_N)$  where  $\Gamma_N = \begin{pmatrix} 2D_p^+(\Sigma \otimes \Sigma)D_p^{+'} & 0 \\ 0 & \Sigma \end{pmatrix}$ , any discrepancy function  $F$

such that  $\frac{1}{2}\ddot{F}(\beta_0) = \Gamma_N^{-1}$  is correctly specified. This includes the ML and GLS approaches

discussed above. As an alternative, estimation can be carried out without making any

---

<sup>1</sup>  $D_p$  is such that  $vecA = D_p vecA$  for any symmetric  $p \times p$  matrix  $A$ , and  $D_p^+$  is its Moore-Penrose inverse.

distributional assumptions and using a completely or partially unstructured estimate of  $\Gamma^{-1}$ .

For example, in the ADF (asymptotically distribution free) approach,

$$\hat{\Gamma} = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})(z_i - \bar{z})', \text{ where } z_i = (x_i', \text{vech}(x_i x_i'))' \text{ and } x_i \text{ is the } i\text{th row of } X. \text{ This}$$

unstructured estimate has the property that  $\hat{\Gamma} \xrightarrow{p} \Gamma$  as long as the distribution of the data has finite fourth-order moments (Browne, 1984). The ADF estimator obtained by minimizing  $F_{ADF} = (\hat{\beta} - \beta(\theta))' \hat{\Gamma}^{-1} (\hat{\beta} - \beta(\theta))$  is asymptotically efficient within the class of estimators based on  $\hat{\beta}$ , and the associated test statistic  $n\hat{F}_{ADF}$  is chi-square distributed. Despite its appealing asymptotic properties, this method has been found to perform extremely poorly unless the sample size is very large (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992), and the first approach of making restrictive assumptions about  $\Gamma$  is often resorted to in practice. More often than not, these assumptions are not true, and the first approach leads to the use of a discrepancy function that is not correctly specified. The literature on asymptotic robustness shows that there exist conditions under which the discrepancy function  $F_{ML}$  is misspecified yet  $\hat{\theta}$  is still asymptotically efficient and  $n\hat{F}_{ML}$  is still chi-square distributed (e.g., Browne, 1987; Mooijart & Bentler, 1991; Satorra & Bentler, 1990; Shapiro, 1987). Because these are also the properties of the ADF estimator, the usual approach to demonstrating asymptotic robustness of a certain estimator and test statistic is to show that under these conditions they are asymptotically equivalent to the ADF estimator and test statistic.

We now define  $C = \Gamma - V^{-1}$ . Following Mooijart and Bentler (1991), we say that  $\Gamma$  satisfies *the robustness condition* if there exists a symmetric  $q \times q$  matrix  $D$  such that  $\Gamma = V^{-1} + \Delta D \Delta'$ , which implies that  $C = \Delta D \Delta'$ . The robustness condition can be equivalently written as  $V = \Gamma^{-1} + \Gamma^{-1} \Delta \bar{D} \Delta' \Gamma^{-1}$ , for some other symmetric matrix  $\bar{D}$  (e.g., Mooijart &

Bentler, 1991). If this condition is satisfied, the estimator  $\hat{\theta}$  obtained by minimizing  $F = (\hat{\beta} - \beta(\theta))' \hat{V} (\hat{\beta} - \beta(\theta))$  for any  $\hat{V} \xrightarrow{p} V$  is asymptotically efficient, and the associated test statistic  $n\hat{F}$  is asymptotically chi-square distributed. The proof relies on the following property of GLS discrepancy functions. The functions  $F_1 = (\hat{\beta} - \beta(\theta))' V (\hat{\beta} - \beta(\theta))$  and  $F_2 = (\hat{\beta} - \beta(\theta))' (V + V\Delta\bar{D}\Delta'V) (\hat{\beta} - \beta(\theta))$ , where  $\bar{D}$  is any symmetric matrix such that  $V + V\Delta\bar{D}\Delta'V$  is positive definite, have the same minimizer and the same minimum (Browne, 1987; Mooijaart & Bentler, 1991; Rao & Mitra, 1971). For NT methods, the robustness condition is given by  $\Gamma = \Gamma_N + \Delta D \Delta'$ . Whenever this equality holds, the NT estimator is asymptotically equivalent to the ADF estimator and thus is asymptotically efficient within the class of estimators based on  $\hat{\beta}$ . Furthermore, the NT test statistic is asymptotically equivalent to the ADF test statistic, and thus is asymptotically chi-square distributed regardless of the distribution of the data.

We now only concern ourselves with the robustness of NT methods, and redefine

$C = \Gamma - \Gamma_N$ . With complete data, we can partition  $\Gamma = \begin{pmatrix} \tilde{\Gamma} & \Gamma_{21}' \\ \Gamma_{21} & \Sigma \end{pmatrix}$ , and correspondingly

$C = \begin{pmatrix} \tilde{C} & \Gamma_{21}' \\ \Gamma_{21} & 0 \end{pmatrix}$ , where  $\tilde{C} = \tilde{\Gamma} - 2D_p^+(\Sigma \otimes \Sigma)D_p^{+'}$  contains population fourth-order cumulants

and  $\Gamma_{21}$  contains population third-order cumulants. We also assume the mean structure is

saturated. This allows us to partition  $\Delta = \begin{pmatrix} \tilde{\Delta} & 0 \\ 0 & I \end{pmatrix}$  and correspondingly  $D = \begin{pmatrix} D_{11} & D_{21}' \\ D_{21} & D_{22} \end{pmatrix}$ . For

the equation  $C = \Delta D \Delta'$  to hold for some  $D$ , it must be that  $D_{22} = 0$ ,  $\tilde{C} = \tilde{\Delta} D_{11} \tilde{\Delta}'$ , and

$\Gamma_{21} = D_{21} \tilde{\Delta}'$ . Robustness properties of NT methods were originally studied in the context of

covariance structures, and hence in most sources the robustness condition is stated in terms of

the conditions on fourth-order moments only; that is,  $\tilde{C} = \tilde{\Delta}D_{11}\tilde{\Delta}'$  (e.g., Browne, 1987; Mooijaart & Bentler, 1991). However, when the mean structure is saturated, whether estimation is based on  $\sigma$  or on  $\beta$  is irrelevant because the corresponding estimates of covariance structure parameters are the same (for GLS discrepancy functions) or asymptotically equivalent (for the ML discrepancy function), even when the data are nonnormal. Thus, considering the structure of  $\tilde{C}$  alone is sufficient. The fact that  $\tilde{C}$  is the matrix of fourth-order cumulants of the observed variables has been used to investigate what types of data and models would lead to asymptotic robustness of NT methods with complete data. Unfortunately, in the incomplete data case, the elements of  $\tilde{C}$  will no longer have an easy interpretation. The incomplete data case is now discussed in detail.

## 2. Asymptotic Robustness with Incomplete Data

With incomplete data, the  $n \times p$  data matrix  $X$  now has empty cells and thus its rows no longer constitute an i.i.d. sample. Rather they can be viewed as drawn from a  $p_j$ -variate distribution  $F_j$  with probability  $q_j$ , where  $j = 1, \dots, 2^p - 1$ . In other words,  $j$  enumerates the missing data patterns. For simplicity, let  $j = 1$  be the index of the complete data pattern, so that  $p_1 = p$ . Then, the following restriction needs to be imposed on the distributions  $F_j$ : for each  $j$ , the mean  $\mu_j$  and the covariance matrix  $\Sigma_j$  must be the appropriate submatrices of  $\mu_1 = \mu$  and  $\Sigma_1 = \Sigma$ . Another way to think about this is to view the rows of  $X$  as drawn from a mixture distribution where the components have different dimensions. As with complete data, we now assume that  $\mu = \mu(\theta)$ ,  $\Sigma = \Sigma(\theta)$ ,  $\theta$  is a  $q \times 1$  vector of parameters,  $\sigma = \text{vech}\Sigma$ , and  $\beta = (\sigma', \mu')'$ .

We obtain  $\hat{\beta}$  by maximizing the normal theory saturated log-likelihood:

$$l(\beta) = \sum_{i=1}^n l_i(\beta) = \frac{1}{2} \sum_{i=1}^n (p_i \log(2\pi) - \log |\Sigma_i(\beta)| - (x_i - \mu_i(\beta))' \Sigma_i^{-1}(\beta) (x_i - \mu_i(\beta))),$$

where  $p_i$  is the dimension of  $x_i$ ,  $E(x_i) = \mu_i$ , and  $\text{cov}(x_i) = \Sigma_i$ . This estimator is analogous to  $\bar{x}$  and  $S$

for complete data. We obtain  $\hat{\theta}$  by maximizing the normal theory structured log-likelihood:

$$l(\theta) = \sum_{i=1}^n l_i(\theta) = \frac{1}{2} \sum_{i=1}^n (p_i \log(2\pi) - \log |\Sigma_i(\theta)| - (x_i - \mu_i(\theta))' \Sigma_i^{-1}(\theta) (x_i - \mu_i(\theta))).$$

This estimator is called direct, raw, or full-information maximum likelihood estimator. As before, the

likelihood ratio test statistic is given by  $T_{ML,inc} = -2(l(\hat{\theta}) - l(\hat{\beta}))$ . By analogy with the

complete data case, we can also define  $\hat{\theta}$  as the minimizer of a “discrepancy function”

$$F_{ML,inc} = \frac{1}{n} \sum_{i=1}^n (\log |\Sigma_i(\theta) \Sigma_i^{-1}(\hat{\beta})| + (x_i - \mu_i(\theta))' \Sigma_i^{-1}(\theta) (x_i - \mu_i(\theta)) - (x_i - \mu_i(\hat{\beta}))' \Sigma_i^{-1}(\hat{\beta}) (x_i - \mu_i(\hat{\beta}))),$$

so that  $T_{ML,inc} = n\hat{F}_{ML,inc} = nF_{ML,inc}(\hat{\theta})$ . Strictly speaking, however, this function does not meet

the definition of a discrepancy function because it also depends on the individual  $x_i$ 's.

Before we proceed, we need to address the consistency of the estimator  $\hat{\beta}$ . This property comes largely for free with complete data (e.g., Shapiro, 1985) but is no longer trivial with incomplete data. In the missing data literature, the following well-known classification of missing data mechanisms is often employed (e.g., Little & Rubin, 2002). Data are said to be *missing completely at random* (MCAR) if the missingness mechanism is independent of both the observed and the missing values of  $X$ ; that is, if

$f(M | X_{obs}, X_{mis}) = f(M)$ , where  $f$  describes the distribution of  $M$ , the indicator matrix of missingness, and  $X_{obs}$  and  $X_{mis}$  represent the observed and the missing part of the data

matrix  $X$ , respectively. Data are said to be *missing at random* (MAR) if the missingness mechanism depends on the observed values but not on the missing values of  $X$ ; that is, if

$f(M | X_{obs}, X_{mis}) = f(M | X_{obs})$ . Otherwise, data are said to be *not missing at random* (NMAR). When data are MCAR,  $\hat{\beta}$  is consistent (Rubin, 1987; Little & Rubin, 2002). When data are MAR and normally distributed,  $\hat{\beta}$  is still consistent (Rubin, 1987; Little & Rubin, 2002). When data are MAR and nonnormally distributed, it is usually said that consistency of  $\hat{\beta}$  cannot be established (e.g., Laird, 1988; Rotnitzky & Wypij, 1994; Yuan & Bentler, 2000), although recently Yuan (2005) argued that  $\hat{\beta}$  remains consistent under MAR mechanism even with nonnormal data. With NMAR data, consistency of  $\hat{\beta}$  or any estimate that does not take the missing mechanism into account cannot be guaranteed. Our working assumption in this article is that data are MCAR, because in addition to consistency of  $\hat{\beta}$  we also require that it have a known asymptotic covariance matrix, as defined below.

Even though the standard central limit theorem no longer applies, alternate methods allow us to assume that  $\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, \Gamma_{inc})$  for some positive definite matrix  $\Gamma_{inc}$  (Yuan & Bentler, 2000). If all  $F_j$ 's are normal, we write instead  $\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, \Gamma_{N,inc})$ . The information matrix for the parameters of the saturated model,  $I = \Gamma_{N,inc}^{-1}$ , now has a more general definition appropriate for incomplete data. Yuan and Bentler (2000) showed that,

when the missing mechanism is MCAR,  $I = \Gamma_{N,inc}^{-1} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \frac{1}{2} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i & 0 \\ 0 & \tau_i' \Sigma_i^{-1} \tau_i \end{pmatrix}$ ,

where  $\kappa_i = D_{p_i} \frac{\partial \text{vech} \Sigma_i}{\partial \sigma'} = \frac{\partial \text{vec} \Sigma_i}{\partial \sigma'}$  and  $\tau_i = \frac{\partial \mu_i}{\partial \mu'}$ .<sup>2</sup> This formula reflects a well-known fact that the information obtained from independent observations is additive. Unfortunately it also makes clear that  $\Gamma_{N,inc}$ , the inverse of this information matrix, no longer has a nice form, and

---

<sup>2</sup> The following identity relates these 0-1 matrices:  $\kappa_i = (\tau_i \otimes \tau_i) D_p$

we can anticipate that  $\Gamma_{inc}$  will not either. When data are MAR, the matrix  $I = \Gamma_{N,inc}^{-1}$  can no longer be guaranteed to be longer block-diagonal (Kenward & Molenberghs, 1998), and furthermore its upper block containing information about the variability of covariance estimates depends on the particular missing mechanism (Yuan, 2005). For this reason, we restrict our attention to MCAR data.

For incomplete data, the information matrix for saturated parameters  $\Gamma_{N,inc}^{-1}$  can be defined as  $\Gamma_{N,inc}^{-1} = -\lim_{n \rightarrow \infty} \frac{1}{n} \ddot{l}(\beta_0)$ , where  $\ddot{l}(\beta) = \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}$ , and hence  $\Gamma_{N,inc} = \lim_{n \rightarrow \infty} n(\ddot{l}(\hat{\beta}))^{-1}$ . The

matrix  $\Gamma_{inc}$  can be defined as  $\Gamma_{inc} = \Gamma_{N,inc} B \Gamma_{N,inc}$ , where  $B = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \dot{l}_i(\hat{\beta}) \dot{l}_i(\hat{\beta})'$  (Arminger & Sobel, 1990; Yuan & Bentler, 2000). With complete data, this result can be verified by working with the saturated log-likelihood. With incomplete data, this triple product expression may be the only way to obtain an equation for the matrix  $\Gamma_{inc}$ . An explicit form

$$\text{for } B \text{ with incomplete data is } B = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \frac{1}{4} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i \tilde{\Gamma} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i & (sym) \\ \frac{1}{2} \tau_i' \Sigma_i^{-1} \tau_i \Gamma_{21} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i & \tau_i' \Sigma_i^{-1} \tau_i \end{pmatrix}.$$

An outline of proof is given in Appendix I.

The cumbersome limit notation can be avoided by grouping the population by missing patterns. With  $p$  variables, there are at most  $k = 2^p - 1$  missing data patterns. As before, let  $q_j$  be the probability that a randomly chosen case in the population has the pattern  $j$ , where  $j = 1, \dots, k$ . We do not have to assume that every pattern exists in the population, and many  $q_j$ 's can be zero. In the sample, the obvious estimate of each  $q_j$  is  $\hat{q}_j = \frac{n_j}{n}$ , where  $n_j$  is the number of cases exhibiting pattern  $j$ . This notation allows us to rewrite

$$\Gamma_{N,inc}^{-1} = \begin{pmatrix} \frac{1}{2} \sum_{j=1}^k q_j \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j & 0 \\ 0 & \sum_{j=1}^k q_j \tau_j' \Sigma_j^{-1} \tau_j \end{pmatrix} \quad (1)$$

$$B = \begin{pmatrix} \frac{1}{4} \sum_{j=1}^k q_j \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j \tilde{\Gamma} \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j & (sym) \\ \frac{1}{2} \sum_{j=1}^k q_j \tau_j' \Sigma_j^{-1} \tau_j \Gamma_{21} \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j & \sum_{j=1}^k q_j \tau_j' \Sigma_j^{-1} \tau_j \end{pmatrix} \quad (2)$$

We can obtain an explicit expression for  $\Gamma_{inc}$  by computing  $\Gamma_{inc} = \Gamma_{N,inc} B \Gamma_{N,inc}$  using (1) and (2).

The approach to robustness reviewed in the previous section applies to GLS function. To extend it to  $F_{ML,inc}$ , we must first show that the minimizer  $\hat{\theta}$  of  $F_{ML,inc}$  is asymptotically equivalent to the minimizer of a quadratic form  $F_{GLS,inc} = (\hat{\beta} - \beta(\theta))' \hat{V} (\hat{\beta} - \beta(\theta))$  for some matrix  $\hat{V} \rightarrow \Gamma_{N,inc}^{-1}$ . This is shown in Appendix II (see also Yuan & Bentler, 2000). It is also shown that  $n\hat{F}_{ML,inc}$  and  $n\hat{F}_{GLS,inc}$  are asymptotically equivalent. Also let

$F_{ADF,inc} = (\hat{\beta} - \beta(\theta))' \hat{\Gamma}_{inc}^{-1} (\hat{\beta} - \beta(\theta))$  be the ADF discrepancy function for incomplete data. By

the argument in the previous section, when the robustness condition  $\Gamma_{inc} = \Gamma_{N,inc} + \Delta D \Delta'$

holds for some symmetric  $q \times q$  matrix  $D$ , the minimizers (and the minima) of  $F_{ML,inc}$  and of

$F_{ADF,inc}$  are asymptotically equivalent. That is, the direct ML estimator  $\hat{\theta}$  is asymptotically

efficient within the class of estimators based on  $\hat{\beta}$ , and the direct ML test statistic  $T_{ML,inc}$  is

asymptotically a chi-square variate with  $p^* - q$  degrees of freedom. As with complete data,

we can define  $C = \Gamma_{inc} - \Gamma_{N,inc}$  and write the robustness condition in the alternative form

$C = \Delta D \Delta'$ . Further, as with complete data, when the mean structure is saturated the

robustness condition simplifies to  $\tilde{C} = \tilde{\Delta} D_1 \tilde{\Delta}'$ , where all the matrices are defined as before.

### 3. From Complete Data to Incomplete Data?

So far, we have shown that the same abstract robustness condition must be met for complete and incomplete data to ensure asymptotic robustness of the NT estimator and associated test statistic. However, the abstract condition  $C = \Delta D \Delta'$  is useless in practice unless some way exists of verifying whether it holds. With complete data, researchers have identified sets of conditions for different models under which this condition holds. As referenced in the introduction, many different types of models have been studied, including CFA models, LISREL models, growth curve models, multiple group models, and multilevel models. One approach, therefore, would be to try to develop similar results for incomplete data. However, as we shall see, this may not be possible. Instead, we address the following question: If we know that the robustness condition is satisfied for a particular complete data population and a particular model, will it still be satisfied for incomplete data obtained from this population by employing an MCAR missing mechanism? This question is equivalent to the following question: if, in the incomplete data population, the robustness condition holds separately for each missing data pattern, does it hold for the overall incomplete data population? Unfortunately, the answer seems to be no.

To see this, it is convenient to rewrite the condition  $\Gamma_{inc} = \Gamma_{N,inc} + \Delta D \Delta'$  in the equivalent form

$$B = \Gamma_{N,inc}^{-1} + \Gamma_{N,inc}^{-1} \Delta D \Delta' \Gamma_{N,inc}^{-1} \quad (3)$$

The matrix  $B$  was defined in (2). This form of the robustness condition avoids the inversion of the information matrix  $\Gamma_{N,inc}^{-1}$  given in (1). Assuming saturated mean structure, the condition in (3) becomes:

$$B = \Gamma_{N,inc}^{-1} + \begin{pmatrix} \frac{1}{4} \sum_{j=1}^k q_j \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j \tilde{\Delta} D_{11} \tilde{\Delta}' \sum_{j=1}^k q_j \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j & (sym) \\ \frac{1}{2} \sum_{j=1}^k q_j \tau_j' \Sigma_j^{-1} \tau_j D_{21} \tilde{\Delta}' \sum_{j=1}^k q_j \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j & 0 \end{pmatrix} \quad (4)$$

When (4) holds for some  $D = \begin{pmatrix} D_{11} & D_{21}' \\ D_{21} & 0 \end{pmatrix}$ , the robustness condition holds in the incomplete data population.

Now, let us assume instead that robustness holds for each missing pattern, to see what this implies about the structure of the overall matrix  $B$ . Note that the expression in (2) does

not involve the individual matrices  $\Gamma_j = \begin{pmatrix} \tilde{\Gamma}_j & \Gamma_{j,21}' \\ \Gamma_{j,21} & \Sigma_j \end{pmatrix}$ . These matrices can be reinserted by

noting the following relationships between the components of the complete data matrix  $\Gamma$

and the components of matrices for individual patterns:  $\tilde{\Gamma}_j = D_{p_j}^+ \kappa_j \tilde{\Gamma} \kappa_j' D_{p_j}^{+'}$  and

$\Gamma_{21,j} = \tau_j \Gamma_{21} \kappa_j' D_{p_j}^{+'}$ . However, this is not necessary, because as these relationships imply, the

condition  $\Gamma = \Gamma_N + \Delta D^* \Delta'$  for some  $q \times q$  symmetric matrix  $D^*$  is equivalent to the condition

$\Gamma_j = \Gamma_{N,j} + \Delta_j D^* \Delta_j'$  for every  $j$ , where  $\Delta_j = \begin{pmatrix} D_{p_j}^+ \kappa_j \tilde{\Delta} & 0 \\ 0 & \tau_j \end{pmatrix}$ . Equivalently, the derivative

matrix for each pattern can be defined as the  $(p_i^* + p_i) \times q$  matrix<sup>3</sup>  $\Delta_j = \frac{\partial \beta_j(\theta)}{\partial \theta'}$ , in general no

longer full rank, where  $\hat{\beta}_j$  be the unstructured ML estimate of  $\beta_j$  based on  $n_j$  observations

in pattern  $j$ . Thus, saying that robustness holds for each pattern is equivalent to saying that it

holds in the complete data population, and the same matrix  $D^*$  works for all patterns.

---

<sup>3</sup> Technically speaking, this matrix should be  $(p_i^* + p_i) \times q_i$ , where  $q_i$  is the length of the subset of  $\theta$  that structures the variables in a given pattern., and  $D^*$  for a particular pattern reduced accordingly. However, as defined,  $\Delta_j$  has columns of zeros corresponding to first and second order moments of variables missing from the pattern, and the corresponding rows and columns of  $D^*$  are irrelevant for that particular pattern.

Using the assumption that robustness holds in the complete data population, and with a little algebra, it can be shown that the structure of  $B$  is given by

$$B = \Gamma_{N,inc}^{-1} + \begin{pmatrix} \frac{1}{4} \sum_{j=1}^k q_j \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j \tilde{\Delta} D_{11}^* \tilde{\Delta}' \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j & (sym) \\ \frac{1}{2} \sum_{j=1}^k q_j \tau_j' \Sigma_j^{-1} \tau_j D_{21}^* \tilde{\Delta}' \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j & 0 \end{pmatrix} \quad (5)$$

It is clear that if  $B$  has the structure given by (5) for some symmetric matrix  $D^*$  (which is true when robustness holds separately for each pattern, or equivalently when it holds for complete data), it does not follow in general that  $B$  can be written in the form given by (4) for some other symmetric matrix  $D$  (which is true when robustness holds for the incomplete data population obtained from complete data).

Let us compare (4) and (5) in more detail. Because we are assuming a saturated mean structure, it is sufficient to consider the  $p^* \times p^*$  upper submatrices of equations (4) and (5).

Let  $T_j = \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j$ . For robustness to hold for the entire incomplete data population,

we must find a matrix  $D_{11}$  such that  $\sum_{j=1}^k q_j T_j \tilde{\Delta} D_{11}^* \tilde{\Delta}' T_j = \sum_{j=1}^k q_j T_j \tilde{\Delta} D_{11} \tilde{\Delta}' \sum_{j=1}^k q_j T_j$ . Such a matrix

only exists when the vector  $\sum_{j=1}^k q_j (T_j \otimes T_j) (\tilde{\Delta} \otimes \tilde{\Delta}) \text{vec} D_{11}^*$  can be expressed as a linear

combination of the columns of the matrix  $(\sum_{j=1}^k q_j T_j) \otimes (\sum_{j=1}^k q_j T_j) (\tilde{\Delta} \otimes \tilde{\Delta})$ . This is not possible,

in general. One rather useless exception is if  $\kappa_j' \kappa_t = 0$  for all  $j \neq t$ , which implies that each variable is observed in only one pattern. In this case, the double summation on the right-hand side of (4) becomes a single summation, leading to (5).

Specific conditions developed to ensure robustness for complete data will no longer ensure robustness if the data become incomplete. Of course,  $\Gamma_{N,inc}^{-1} \rightarrow \Gamma_N^{-1}$  and  $\Gamma_{inc}^{-1} \rightarrow \Gamma^{-1}$  as the proportion of complete cases approaches 1, so that the extent to which robustness is

violated when the data are no longer complete will depend on the proportion of missing data. Additional factors, such as the extent of nonnormality of the data, may also exert an influence. Some examples are given in the next section.

#### 4. Illustrations

##### *Example 1*

We first illustrate how robustness is violated when going from complete data to incomplete data on a very simple example.<sup>4</sup> Let  $x_1$  and  $x_2$  be mutually independent random variables with fourth-order cumulants equal to 1 and the covariance matrix  $\Sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$ . The saturated estimate of the covariance matrix will contain three nonredundant elements, and therefore this model has one degree of freedom. We do not care about the means and omit the “tilde” notation from all the matrices for convenience. The derivative matrix for this model is

$$\Delta = \frac{\partial \text{vech}\Sigma}{\partial \theta'} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \text{ where } \theta = (\sigma_{11}, \sigma_{22})'. \text{ We have}$$

$$\Gamma_N = 2D_2^+(\Sigma \otimes \Sigma)D_2^{+'} = \begin{pmatrix} 2\sigma_{11}^2 & 0 & 0 \\ 0 & \sigma_{11}\sigma_{22} & 0 \\ 0 & 0 & 2\sigma_{22}^2 \end{pmatrix}, \text{ and the cumulants matrix is}$$

$$C = \Gamma - \Gamma_N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \text{ because the variables are mutually independent. This gives us } \Gamma$$

in the complete data population. Robustness condition  $C = \Delta D \Delta'$  holds with  $D = I_2$ .

We now create incomplete data by deleting the value of  $x_2$  from every other row of the data. This process generates MCAR data with two patterns, and the probabilities of each

---

<sup>4</sup> I would like to thank Bob Jennrich for suggesting this example.

pattern are  $q_1 = q_2 = .5$ . In pattern 1, all the matrices are defined as for complete data. In

pattern 2,  $\Gamma_{2,N} = 2(\sigma_{11} \otimes \sigma_{11}) = 2\sigma_{11}^2$  and  $\Gamma_2 = 2\sigma_{11}^2 + 1$ . For the resulting incomplete data, we

have:

$$\begin{aligned} \Gamma_{inc,N} &= \frac{1}{2} \left( .5\kappa_1'(\Sigma_1^{-1} \otimes \Sigma_1^{-1})\kappa_1 + .5\kappa_1'(\Sigma_2^{-1} \otimes \Sigma_2^{-1})\kappa_2 \right)^{-1} \\ \Gamma_{inc,N} &= \left( \begin{pmatrix} 1/2\sigma_{11}^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1/2\sigma_{11}^2 & 0 & 0 \\ 0 & 1/\sigma_{11}\sigma_{22} & 0 \\ 0 & 0 & 1/2\sigma_{22}^2 \end{pmatrix} \right)^{-1} \\ \Gamma_{inc,N} &= \begin{pmatrix} \sigma_{11}^2 & 0 & 0 \\ 0 & \sigma_{11}\sigma_{22} & 0 \\ 0 & 0 & 2\sigma_{22}^2 \end{pmatrix} \end{aligned} \quad (6)$$

Similarly,

$$\begin{aligned} B &= \frac{1}{4} (.5\kappa_1'(\Sigma_1^{-1} \otimes \Sigma_1^{-1})\kappa_1 \Gamma \kappa_1'(\Sigma_1^{-1} \otimes \Sigma_1^{-1})\kappa_1 + .5\kappa_2'(\Sigma_2^{-1} \otimes \Sigma_2^{-1})\kappa_2 \Gamma \kappa_2'(\Sigma_2^{-1} \otimes \Sigma_2^{-1})\kappa_2) \\ B &= \begin{pmatrix} \frac{(2\sigma_{11}^2+1)}{4\sigma_{11}^4} & 0 & 0 \\ 0 & 1/2\sigma_{11}\sigma_{22} & 0 \\ 0 & 0 & \frac{(2\sigma_{22}^2+1)}{8\sigma_2^4} \end{pmatrix} \end{aligned} \quad (7)$$

Using (6) and (7) in  $\Gamma_{inc} = \Gamma_{inc,N} B \Gamma_{inc,N}$ , we obtain

$$\Gamma_{inc} = \begin{pmatrix} \frac{(2\sigma_{11}^2+1)}{4} & 0 & 0 \\ 0 & \sigma_{11}\sigma_{22}/2 & 0 \\ 0 & 0 & \frac{(2\sigma_{22}^2+1)}{8} \end{pmatrix} \quad (8)$$

Using (6) and (8) in  $C = \Gamma_{inc} - \Gamma_{inc,N}$ , we obtain

$$C = \begin{pmatrix} \frac{1-2\sigma_{11}^2}{4} & 0 & 0 \\ 0 & -\frac{\sigma_{11}\sigma_{22}}{2} & 0 \\ 0 & 0 & \frac{1-2\sigma_{22}^2}{4} \end{pmatrix}$$

Clearly, robustness condition is not satisfied for this matrix because it has rank 3, whereas  $\Delta$  has rank 2. For completeness, we write out

$$\Delta D \Delta' = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d_{11} & 0 \\ 0 & d_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} d_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & d_{22} \end{pmatrix} \neq C \text{ for any } d_{11}, d_{22}.$$

### Example 2

We focus on the standard CFA model as our second example. The violation of robustness when going from complete data from incomplete data will be illustrated empirically as well as mathematically. The CFA model is usually stated as  $x = \mu + \Lambda\xi + \zeta$ , where  $x$  is a  $p \times 1$  vector of observed variables,  $\xi$  is an  $m \times 1$  vector of latent factors,  $\zeta$  is a  $p \times 1$  vector of errors,  $\mu$  is a  $p \times 1$  vector of constants, and  $\Lambda$  is a  $p \times m$  matrix of factor loadings. We assume that  $E(\xi) = 0$  and  $E(\zeta) = 0$ , so that  $E(x) = \mu$ . We also assume that  $E(\xi\xi') = 0$  and  $E(\zeta_i\zeta_j) = 0$  for all  $i \neq j$ , so that  $\Sigma = \Lambda\Phi\Lambda' + \Psi$ , where  $E(\xi\xi') = \Phi$ ,  $E(\zeta\zeta') = \Psi$  (a diagonal matrix), and  $E(xx') - \mu\mu' = \Sigma$ . The asymptotic robustness of the NT methods under the CFA model has been studied extensively (Amemiya & Anderson, 1990; Anderson & Amemiya, 1988; Browne & Shapiro, 1988; Mooijart & Bentler, 1991; Satorra & Bentler, 1990). It has been shown that the robustness condition is satisfied in this case when all of the following conditions hold (assuming complete data):

- i. The variables  $\xi$  are independent of the variables  $\zeta$ ;

- ii. The variables  $\zeta$  are mutually independent (or, if a block of them forms a dependent set, the variables in this block have no excess kurtosis and are independent of the variables not in the block, and the covariances of the variables within the block must be freely estimated);
- iii. All elements of  $\Phi$  are freely estimated (or, if some elements of  $\Phi$  are fixed to zero, the corresponding factors must be independent).

Because the CFA model does not structure the means, skewness of any of the variables involved will have no effect on robustness. We also note that any model that can be equivalently parameterized as a model satisfying these conditions will satisfy the robustness condition. For example, it is common to fix the variance of a factor to 1, which contradicts condition (iii); however, there exists an equivalent factor model where one of the loadings is fixed to 1 instead.

We consider a very small factor model as an example. This model has 1 factor and three indicators, with all the factor loadings fixed to 1. Conditions (i)-(iii) above do not prohibit restrictions on the elements of  $\Lambda$ . Because the means are unstructured we again drop the “tilde” notation on all the relevant matrices for simplicity. This model implies the following system of equations:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \xi + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix} \quad (9)$$

and the following covariance structure:

$$\begin{pmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{31} \\ \sigma_{22} \\ \sigma_{32} \\ \sigma_{33} \end{pmatrix} = \begin{pmatrix} \phi_{11} + \psi_{11} \\ \phi_{11} \\ \phi_{11} \\ \phi_{11} + \psi_{22} \\ \phi_{11} \\ \phi_{11} + \psi_{33} \end{pmatrix} \quad (10)$$



where  $k_\xi = E(\xi_1^4) - 3E^2(\xi_1^2)$  and  $k_i = E(\varepsilon_i^4) - 3E^2(\varepsilon_i^2)$ . That is, the matrix of cumulants of the observed variables can be written in terms of cumulants of the latent variables because the latent variables are mutually independent (Browne, 1987; Mooijaart and Bentler, 1991).

When this is the case, the matrix

$$D^* = \begin{pmatrix} k_\xi & 0 & 0 & 0 \\ 0 & k_1 & 0 & 0 \\ 0 & 0 & k_2 & 0 \\ 0 & 0 & 0 & k_3 \end{pmatrix} \quad (13)$$

satisfies the equation  $C = \Delta D^* \Delta$ , where  $\Delta$  and  $C$  are given by (11) and (12). We have shown that robustness holds for this small model under assumptions (i)-(iii).

Now assume that the population is a mixture of two patterns: 25% of the cases have all three variables observed, and 75% of the cases have only the first two variables observed. That is, any randomly drawn vector from this population has .25 probability of being from model given by (9) and (10), and a .75 probability of being from the model

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \xi + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \text{ with covariance structure } \begin{pmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{22} \end{pmatrix} = \begin{pmatrix} \phi_{11} + \psi_{11} \\ \phi_{11} \\ \phi_{11} + \psi_{22} \end{pmatrix}. \text{ Combining the two}$$

patterns into one population, we obtain  $\Gamma_{inc,N} = \frac{1}{2}(.25T_1 + .75T_2)^{-1}$ , where

$T_j = \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j$ . It is clear that  $\Gamma_{inc,N}$  no longer has a simple expression in terms of the elements of  $\Sigma$ , even in this small example. The complication in general arises from the fact that the elements of  $\Sigma_j^{-1}$  are very complicated functions of elements of  $\Sigma$ . We also obtain

$B = \frac{1}{16}(T_1 \Gamma T_1 + 3T_2 \Gamma T_2)$ . We consider the robustness condition in the equivalent form given

by (3). We have that  $B - \Gamma_{inc,N}^{-1} = \frac{1}{16}(1T_1 \Delta D^* \Delta' T_1 + 3T_2 \Delta D^* \Delta' T_2)$  whereas

$\Gamma_{inc,N}^{-1} \Delta D \Delta \Gamma_{inc,N}^{-1} = \frac{1}{16^2} (T_1 + 3T_2) \Delta D \Delta (T_1 + 3T_2)$ , and a solution for  $D$  exists only if the vector

$(T_1 \otimes T_1 + 3T_2 \otimes T_2)(\Delta \otimes \Delta) \text{vec} D^*$  is in the column space of the matrix

$\frac{1}{16} ((T_1 + 3T_2) \otimes (T_1 + 3T_2))(\Delta \otimes \Delta)$ , which is not the case in general. Specific cases can be

entertained by choosing numerical values for  $\Sigma$  and  $\Gamma$ .

Thus, the fact that conditions (i)-(iii) hold for each pattern does not ensure robustness for the incomplete data in this CFA example. There may well exist other conditions that ensure that robustness condition holds, but determining what they are is no longer a straightforward task. With incomplete data, we may not be able to convert the abstract robustness condition into properties of specific models.

To empirically illustrate the influence of missing data on the performance of the ML test statistic as applied to this small model, a simulation study was run in EQS 6.1 (Bentler, 2005). The errors and factors were specified to have zero univariate skewness and varying degrees of univariate kurtosis (0, 10, or 20). EQS uses Fleishman's power method (Fleishman, 1978; Vale & Maurelli, 1983) to create the desired univariate characteristics. An additional condition where the latent variables were specified to have univariate skewness of 3 and univariate kurtosis of 20 was added to illustrate that skewness does not affect robustness for this model. The resulting univariate kurtosis range and average multivariate kurtosis for the observed variables in the four conditions are given in Table 1. Incomplete data were generated from complete data for each nonnormality condition by performing random deletion of a certain percent of observations (0%, 10%, 15%, or 30%). With three variables, this mechanism can result in up to 7 possible missing data patterns. The sample size was set to 1000, which was thought to be high enough to study asymptotic properties of the NT statistic, given the small size of the model. Five hundred replications of each condition were run. The acceptance rates of the ML test statistic are given in Table 1.

It is clear from Table 1 that the ML test statistic is robust with complete data, no matter the degree of nonnormality, but it is no longer robust when the data become incomplete. The degree to which the statistic's performance deviates from expected is a function of both the percent of observations missing and also of the degree of nonnormality of the data. For example, when 10% of the observations are missing and nonnormality is moderate, the ML test statistic rejects 8% of all models instead of the expected 5%--a deviation that can be argued is not very serious. At the other extreme, when 30% of the observations are missing and nonnormality is high, the ML test statistic rejects 21% of all correct models, which is not acceptable performance. Thus, this simulation illustrates that models for which robustness has been shown to hold for incomplete data can no longer be assumed to generate robust NT statistics when the data are incomplete, but the extent to which missing data is a problem can vary.

## 5. Discussion

Literature on the asymptotic robustness of NT methods has demonstrated that NT estimators and test statistics often remain robust to nonnormality for many common structural equation models. For example, in the standard one-factor model, one need only assume that the errors are mutually independent and are independent of the latent factor to ensure robustness. Such an assumption is not at all unreasonable from the point of view of theory, as errors are often conceptualized as random noise. That this popular model is robust to nonnormality under these minimal assumptions is useful information when software that can do other types of estimation is not available. In this paper it was shown, however, that the conditions ensuring robustness for complete data do not generalize to incomplete data.

If proportion of missing data is small and nonnormality is not severe, the extent to which robustness is violated can be ignored. But, as simulation results show, with increasing

proportion of missing data and high nonnormality of the variables, the behavior of the NT statistic deviates from expected by a significant amount. Because missing data is ubiquitous in the social sciences, this finding presents problems for asymptotic robustness theory. At worst, it may be that the theory is best left as a curious theoretical development that has no place in SEM practice. At best, researchers are encouraged not to trust NT test statistics when their data are severely nonnormal and the proportion of missing data is high (at or higher than 15% in this study), even if they have made reasonable assumptions of independence such as those stated above. Better options exist for dealing with nonnormality when the data are incomplete, such as Yuan-Bentler scaled chi-square (Yuan & Bentler, 2000). Recently developed F statistics (Bentler & Yuan 1999b; Yuan & Bentler, 1998a) are especially appropriate for smaller sample sizes; however, their incomplete data versions have not been studied empirically, although they are available in EQS 6.1.

However, the abstract condition  $C = \Delta D \Delta'$  still ensures robustness for both complete and incomplete data. It may be possible to empirically evaluate whether this condition holds by using sample estimates of the matrices involved. If so, an empirical test of asymptotic robustness regardless of the type of data and model may be possible. This line of research will be further pursued.

An analogy is often drawn between modeling with incomplete data and modeling with multiple groups, because incomplete data can be conceptualized as arising from a multiple group population where constraints on all common parameters are imposed across groups. Asymptotic robustness with multiple groups has been investigated (e.g., Satorra, 2001, 2002). It is therefore reasonable to ask whether these results can be used to draw conclusions about robustness with incomplete data. However, there are important differences between these two modeling scenarios. With multiple groups, the overall test statistic tests the proposed model and the equality constraints across groups against the null model where neither the model nor

the constraints are assumed to hold. With incomplete data, the test statistic tests the proposed model with equality constraints imposed on all covariance matrices (here, groups are missing data patterns) against the null model under which the model is not assumed to hold but the constraints still hold. In other words, the equality constraints are tested in the multiple group situation, but they are not tested in the incomplete data situation. The equality constraints on parameters such as factor variances and covariances have been shown to lead to violation of robustness in the multiple group situation (Satorra, 2001), but these results have no bearing on what happens with the NT test statistic with incomplete data. As an example, suppose the data are incomplete, with two missing patterns, and we fit a one-factor model with three indicators. This model is saturated, and hence robustness holds trivially even with incomplete data. But viewed as a two-group model with all parameters constrained across groups, we now have a model with 6 degrees of freedom, and because factor variances are constrained to equal across groups, the resulting test statistic will not be robust, according to Satorra (2001). The correct analogy would be between the multiple group approach and a test statistic for incomplete data that combines a test of model fit with the test of the MCAR assumption (e.g., Kim & Bentler, 2002, and references therein).

## Appendix I

We want to show that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \dot{l}_i(\hat{\beta}) \dot{l}_i(\hat{\beta})' = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left( \begin{array}{cc} \frac{1}{4} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i \tilde{\Gamma} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i & (sym) \\ \frac{1}{2} \tau_i' \Sigma_i^{-1} \tau_i \Gamma_{21} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i & \tau_i' \Sigma_i^{-1} \tau_i \end{array} \right)$$

$$\text{Note that } \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \dot{l}_i(\hat{\beta}) \dot{l}_i(\hat{\beta})' = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E(\dot{l}_i(\hat{\beta}) \dot{l}_i(\hat{\beta})').$$

Recall that the saturated log-likelihood is given by

$$l_i(\beta) = \frac{1}{2} p_i \log(2\pi) - \frac{1}{2} \log |\Sigma_i(\beta)| - \frac{1}{2} (x_i - \mu_i(\beta))' \Sigma_i^{-1}(\beta) (x_i - \mu_i(\beta)).$$

$$\text{Then, } \frac{\partial l_i(\beta)}{\partial (\text{vec} \Sigma_i', \mu_i')} = \left( -\frac{1}{2} (\text{vec} \Sigma_i^{-1})' + \frac{1}{2} (x_i - \mu_i)' \otimes (x_i - \mu_i)' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}), (x_i - \mu_i)' \Sigma_i^{-1} \right).$$

Let  $\beta_i = (\text{vech} \Sigma_i', \mu_i')'$ . From the chain rule, we have the identities:

$$\frac{\partial l_i(\beta)}{\partial \beta_i'} = \frac{\partial l_i(\beta)}{\partial (\text{vec} \Sigma_i', \mu_i')} \begin{pmatrix} D_{p_i} & 0 \\ 0 & I \end{pmatrix}, \text{ and } \frac{\partial l_i(\beta)}{\partial \beta'} = \frac{\partial l_i(\beta)}{\partial \beta_i'} \frac{\partial \beta_i}{\partial \beta'}.$$
 Thus we obtain

$$\frac{\partial l_i(\beta)}{\partial \beta'} = \left( -\frac{1}{2} (\text{vec} \Sigma_i^{-1})' \kappa_i + \frac{1}{2} (x_i - \mu_i)' \otimes (x_i - \mu_i)' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i, (x_i - \mu_i)' \Sigma_i^{-1} \tau_i \right).$$

Define  $\tilde{\Gamma}_i = D_{p_i}^+ (E((x_i - \mu_i) \otimes (x_i - \mu_i)(x_i - \mu_i)' \otimes (x_i - \mu_i)')) - \Sigma_i \otimes \Sigma_i) D_{p_i}^{+'}$  and

$\Gamma_{21,i} = E((x_i - \mu_i)((x_i - \mu_i)' \otimes (x_i - \mu_i)')) D_{p_i}^{+'}$ . Note that these can be related to the complete

data  $\Gamma$  matrix via the identities:  $\tilde{\Gamma}_i = D_{p_i}^+ \kappa_i \tilde{\Gamma} \kappa_i' D_{p_i}^{+'}$  and  $\Gamma_{21,i} = \tau_i \Gamma_{21} \kappa_i' D_{p_i}^{+'}$ .

Then,

$$E \left( \frac{\partial l_i(\beta)}{\partial \sigma} \frac{\partial l_i(\beta)}{\partial \sigma'} \right) = \frac{1}{4} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i \tilde{\Gamma} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i$$

$$E \left( \frac{\partial l_i(\beta)}{\partial \mu} \frac{\partial l_i(\beta)}{\partial \mu'} \right) = \tau_i' \Sigma_i^{-1} \tau_i$$

$$E \left( \frac{\partial l_i(\beta)}{\partial \mu} \frac{\partial l_i(\beta)}{\partial \sigma'} \right) = \frac{1}{2} \tau_i' \Sigma_i^{-1} \tau_i \Gamma_{21} \kappa_i' (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) \kappa_i.$$

## Appendix II

Let  $\hat{\theta}$  be the minimizer of  $F_{ML} = \frac{2}{n}(l(\hat{\beta}) - l(\theta))$  and let  $\tilde{\theta}$  be the minimizer of

$$F_{GLS} = (\hat{\beta} - \beta(\theta))' \hat{V} (\hat{\beta} - \beta(\theta)), \text{ where } \hat{\beta} \xrightarrow{p} \beta_0 = \beta(\theta_0) \text{ and } \hat{V} \xrightarrow{p} V = -\lim_{n \rightarrow \infty} \frac{1}{n} \ddot{l}(\beta_0).$$

Writing  $\beta(\tilde{\theta}) = \beta_0 + \Delta(\tilde{\theta} - \theta_0) + o_p(1/\sqrt{n})$ , we get

$$\sqrt{n}(\hat{\beta} - \beta_0) - \sqrt{n}(\hat{\beta} - \beta(\tilde{\theta})) = \sqrt{n}\Delta(\tilde{\theta} - \theta_0) + o_p(1) \quad (\text{A1})$$

Because  $\tilde{\theta}$  is such that  $\Delta'V(\hat{\beta} - \beta(\tilde{\theta})) \xrightarrow{p} 0$ , pre-multiplying by  $\Delta'V$  yields

$$\sqrt{n}\Delta'V(\hat{\beta} - \beta_0) = \sqrt{n}\Delta'V\Delta(\tilde{\theta} - \theta_0) + o_p(1). \text{ Then,}$$

$$\sqrt{n}(\tilde{\theta} - \theta_0) = \sqrt{n}(\Delta'V\Delta)^{-1} \Delta'V(\hat{\beta} - \beta_0) + o_p(1) \quad (\text{A2})$$

From (A1) and (A2) we get that

$$\sqrt{n}(\hat{\beta} - \beta(\tilde{\theta})) = \sqrt{n}(I - \Delta(\Delta'V\Delta)^{-1} \Delta'V)(\hat{\beta} - \beta_0) + o_p(1) \quad (\text{A3})$$

Using (A3), the test statistic  $T_{GLS} = n(\hat{\beta} - \beta(\tilde{\theta}))' \hat{V} (\hat{\beta} - \beta(\tilde{\theta}))$  can be rewritten as

$$T_{GLS} = n(\hat{\beta} - \beta_0)'(V - V\Delta(\Delta'V\Delta)^{-1} \Delta'V)(\hat{\beta} - \beta_0) + o_p(1) \quad (\text{A4})$$

To obtain the same results for the ML estimator and test statistic, write

$$\frac{1}{n} \dot{l}(\hat{\beta}) = \frac{1}{n} \dot{l}(\beta_0) + \frac{1}{n} \ddot{l}(\beta_0)(\hat{\beta} - \beta_0) + o_p(1/\sqrt{n}) = 0, \text{ where } \dot{l}(\beta) = \frac{\partial l(\beta)}{\partial \beta'} = \sum_{i=1}^n \frac{\partial l_i(\beta)}{\partial \beta} \text{ and}$$

$$\ddot{l}(\beta) = \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}. \text{ Then,}$$

$$\sqrt{n}(\hat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} V^{-1} \dot{l}(\beta_0) + o_p(1) \quad (\text{A5})$$

Similarly, and by the chain rule,

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} (\Delta'V\Delta)^{-1} \Delta' \dot{l}(\beta_0) + o_p(1) \quad (\text{A6})$$

From (A5) and (A6) it follows that

$$\sqrt{n}(\hat{\theta} - \theta_0) = \sqrt{n}(\Delta'V\Delta)^{-1} \Delta'V(\hat{\beta} - \beta_0) + o_p(1) \quad (\text{A7})$$

For the ML test statistic, we need the following expansions:

$$\frac{1}{n}l(\hat{\beta}) = \frac{1}{n}l(\beta_0) + \frac{1}{n}(i(\beta_0))'(\hat{\beta} - \beta_0) + \frac{1}{2n}(\hat{\beta} - \beta_0)'i(\beta_0)(\hat{\beta} - \beta_0) + o_p(1/n) \quad (\text{A8})$$

$$\frac{1}{n}l(\hat{\theta}) = \frac{1}{n}l(\theta_0) + \frac{1}{n}(i(\theta_0))'(\hat{\theta} - \theta_0) - \frac{1}{2}(\hat{\theta} - \theta_0)'\Delta'V\Delta(\hat{\theta} - \theta_0) + o_p(1/n) \quad (\text{A9})$$

Using (A5)-(A7), we can rewrite (A8) and (A9) as follows:

$$\frac{1}{n}l(\hat{\beta}) = \frac{1}{n}l(\beta_0) + \frac{1}{2}(\hat{\beta} - \beta_0)'V(\hat{\beta} - \beta_0) + o_p(1/n) \quad (\text{A10})$$

$$\frac{1}{n}l(\hat{\theta}) = \frac{1}{n}l(\theta_0) + \frac{1}{2}(\hat{\beta} - \beta_0)'V\Delta(\Delta'V\Delta)^{-1} \Delta'V(\hat{\beta} - \beta_0) + o_p(1/n) \quad (\text{A11})$$

Then, under the assumption that  $l(\beta_0) = l(\theta_0)$ ,

$$T_{ML} = 2(l(\hat{\beta}) - l(\hat{\theta})) = n(\hat{\beta} - \beta_0)'(V - V\Delta(\Delta'V\Delta)^{-1} \Delta'V)(\hat{\beta} - \beta_0) + o_p(1) \quad (\text{A12})$$

From (A2) and (A7), we have that  $\hat{\theta}$  and  $\tilde{\theta}$  are asymptotically equivalent. From (A4) and

(A12), we have that  $T_{GLS}$  and  $T_{ML}$  are asymptotically equivalent.

Further, under the assumption that  $\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, \Gamma)$ , we have that

$$\sqrt{n}(\tilde{\theta} - \theta_0) \approx \sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, (\Delta'V\Delta)^{-1} \Delta V \Gamma V \Delta' (\Delta'V\Delta)^{-1}).$$

## References

- Amemiya, Y., & Anderson, T.W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, *18*, 1453-1463.
- Amemiya, Y., Fuller, W.A., & Anderson, T.W. (1987). The asymptotic distributions of some estimators for a factor analytic model. *Journal of Multivariate Analysis*, *22*, 51-64.
- Anderson, T.W., & Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, *16*, 759-771.
- Arminger, G. & Sobel, M. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association*, *85*, 195-203.
- Bentler, P.M. (2005). EQS 6 [Computer software]. Encino, CA: Multivariate Software, Inc.
- Bentler, P.M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: test statistics. *Multivariate Behavioral Research*, *34*, 181-197.
- Bentler, P.M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, *34*, 181-197.
- Browne, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, *8*, 1-24.
- Browne, M.W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62-83.
- Browne, M.W. (1987). Robustness of statistic inference in factor analysis and related models. *Biometrika*, *74*, 375-384.
- Browne, M.W., & Arminger, G. (1995). Specification and estimation of mean and covariance structure models. Pp. 185-249 in *Handbook of Statistical Modeling for the Social and*

- Behavioral Sciences, edited by G.Arminger, C.C.Clogg, and M.E. Sobel. New York: Plenum.
- Browne, M.W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, *41*, 193-208.
- Browne, M.W. (1990). Asymptotic robustness of normal theory methods for the analysis of latent curves. In W. Fuller and P. Brown (Eds.). *Statistical Analysis of Measurement Error Models and Applications*, pp.211-225, American Mathematical Society: Contemporary Mathematics Series, 112.
- Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29.
- Fleishman, A.I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.
- Hu, L., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351-362.
- Kenward, M.G., & Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, *13*, 236-247.
- Kim, K.H., & Bentler, P.M. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika*, *67*, 609-624.
- Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, *7*, 305-315.
- Little, R.J.A., & Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: Wiley.
- Magnus, J.R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics* (revised edition). Chichester UK: Wiley.

- Mooijart, A., & Bentler, P.M. (1991). Robustness of normal theory test statistics in structural equation models. *Statistica Neerlandica*, 45, 159-171.
- Rao, C.R., & Mitra, S.K. (1971). *Generalized Inverse of Matrices and its Applications*. New York: Wiley.
- Rotnitzky, A., & Wypij, D. (1994). A note on the bias of estimators with missing data. *Biometrics*, 50, 1163-1170.
- Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, 54, 131-151.
- Satorra, A. (2001). Goodness of fit testing of structural equation models with multiple group data and nonnormality. In *Structural Equation Modeling: Present and Future*. Cudeck, R., du Toit, S.H.C., & Sörbom, D. (Eds.). pp. 231-258. Lincolnwood, IL: Scientific Software International.
- Satorra, A. (2002). Asymptotic robustness in multiple group linear-latent variable models. *Econometric Theory*, 18, 297-312.
- Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Satorra, A., & Bentler, P.M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis*, 10, 235-249.
- Shapiro, A. (1985). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. *South African Statistical Journal*, 17, 33-81.
- Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association*, 81, 142-149.

- Shapiro, A. (1987). Robustness properties of the MDF analysis of moment structures. *South African Statistical Journal*, 21, 39-62.
- Vale, S.G., & Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48, 465-471.
- Yuan., K.-H. (2005). *Normal theory ML for missing data with violation of distribution assumption*. Unpublished manuscript. University of Notre Dame.
- Yuan, K.-H., & Bentler, P.M.(1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, 51, 289-309.
- Yuan, K.-H., & Bentler, P.M.(2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 165-200.
- Yuan, K.-H., & Bentler, P.M. (2005). Asymptotic robustness of the normal theory likelihood ratio statistic for two-level covariance structure models. *Journal of Multivariate Analysis*, 94, 328-343.

Table 1. Acceptance Rates of the ML test statistic. Cells in italics represent conditions of interest.

Factor & Errors	Univariate Kurtosis Range	Multivariate Kurtosis*	0% Missing	10% Missing	15% Missing	30% Missing
Skew 0; Kurt 0			473/500 (94.8%)	480/500 (96.0%)	477/500 (95.4%)	471/500 (94.2%)
Skew 0; Kurt 10	About 2-6	About 18	468/500 (93.6%)	<i>460/500 (92.0%)</i>	<i>458/500 (91.6%)</i>	<i>426/500 (85.2%)</i>
Skew 0; Kurt 20	About 5-11	About 24	471/500 (94.2%)	<i>442/500 (88.4%)</i>	<i>436/500 (87.2%)</i>	<i>395/500 (79.0%)</i>
Skew 3; Kurt 20	About 5-11	About 36	474/500 (94.8%)	<i>439/500 (87.8%)</i>	<i>437/500 (87.4%)</i>	<i>397/500 (79.4%)</i>

\*Unnormalized estimate is given to yield a sample size independent measure