

## THE NONSINGULARITY OF $\Gamma$ IN COVARIANCE STRUCTURE ANALYSIS OF NONNORMAL DATA

ROBERT JENNRICH

UNIVERSITY OF CALIFORNIA, LOS ANGELES

ALBERT SATORRA

UNIVERSITAT POMPEU FABRA, BARCELONA

Covariance structure analysis of nonnormal data is important because in practice all data are non-normal. When applying covariance structure analysis to nonnormal data, it is generally assumed that the asymptotic covariance matrix  $\Gamma$  for the nonredundant terms in the sample covariance matrix  $S$  is nonsingular. It is shown this need not be the case, which raises a question of how restrictive this assumption may be and how difficult it may be to verify it. It is shown that  $\Gamma$  is nonsingular whenever sampling is from a nonsingular distribution, including any distribution defined by a density function. In the discrete case necessary and sufficient conditions are given for the nonsingularity of  $\Gamma$ , and it is shown how to demonstrate  $\Gamma$  is nonsingular with high probability. Thus, the nonsingularity of  $\Gamma$  assumption is mild and one should feel comfortable about making it. These observations also apply to the asymptotic covariance matrix  $\Gamma$  that arises in structural equation modeling.

Key words: covariance structure analysis, nonredundant components, non-singularity of Gamma, nonsingular distribution, discrete distribution, multivariate polynomials, smooth manifolds.

### 1. Introduction

A basic assumption in covariance structure analysis (CSA) is

$$\sqrt{n}(s - \sigma) \xrightarrow{D} N(0, \Gamma),$$

where  $s = \text{vech}(S)$ ,  $\sigma = \text{vech}(\Sigma)$ , and  $S$  and  $\Sigma$  are sample and population covariance matrices, respectively.

We are interested in the almost universal assumption that  $\Gamma$  is nonsingular. See, for example, Bentler (1983, p. 503, line 4), Browne (1984, R1), Satorra (1989, Theorem 5.2), and Yuan and Bentler (2001, C5). While this assumption has been used for over 30 years, there has been no discussion of its importance. Is this an assumption that almost always holds, or perhaps one that almost never holds? We will show that  $\Gamma$  can be singular when  $\Sigma$  is nonsingular, thus demonstrating that this is a real assumption.

Statistical software programs for CSA are based on theorems that assume  $\Gamma$  is nonsingular. The validity of standard errors and tests produced by these programs is predicated on the validity of this assumption. We show that when this assumption does not hold these programs can fail dramatically and may give no indication that a failure has occurred.

It is the user of these programs who must make this assumption. Thus, if I perform a confirmatory factor analysis using SAS, I'm the one who must assume  $\Gamma$  is nonsingular. Because SAS's confirmatory factor analysis program has many users, one might automatically come to

The research of the second author is supported by grant EC02011-28875 from the Spanish Ministry of Science and Innovation.

Requests for reprints should be sent to Robert Jennrich, University of California, Los Angeles, 3400 Purdue Avenue, Los Angeles, CA, USA. E-mail: [rij@stat.ucla.edu](mailto:rij@stat.ucla.edu)

the conclusion that this assumption must be OK. This is the fifty million Frenchmen argument. Fortunately,  $\Gamma$  is always, or at least almost always, nonsingular. Showing this is the purpose of this paper. We wish to move something that is a matter of faith to a matter of fact.

The assumption that  $\Gamma$  is nonsingular is much more difficult to make than the common assumption that  $\Sigma$  is nonsingular. In regression analysis, for example, one tentatively assumes  $\Sigma$  is nonsingular. If it is not, the sample covariance matrix  $S$  must be singular and one's program reports this. The user generally responds by using a subset of predictors for which  $S$  is nonsingular.

In CSA one's computer cannot in general provide a red flag to warn the user that  $\Gamma$  may be singular. As we will show, when  $\Gamma$  is singular its estimate  $\hat{\Gamma}$  is generally nonsingular. As a consequence, one's computer cannot detect the singularity of  $\Gamma$  from that of its estimate  $\hat{\Gamma}$ .

We begin by showing  $\Gamma$  is always nonsingular when sampling from a nonsingular distribution, including distributions defined by density functions. In this case,  $\Gamma$  is nonsingular and no additional assumptions are required.

From the point of view of applications, the primary singular distributions are the discrete distributions. In this case, we give necessary and sufficient conditions for the nonsingularity of  $\Gamma$  and show how to use these to show  $\Gamma$  is nonsingular in specific applications, at least with high probability. Hopefully, these results will make users of CSA software much more comfortable about the non-singularity assumption on which their software is based.

A reviewer suggested that we show what may happen when  $\Gamma$  is singular. In the [Appendix](#), we show that an extensively used goodness of fit test statistic breaks down completely.

The asymptotic covariance matrix  $\Gamma$  also arises in structural equation modeling, where it is also assumed to be nonsingular. Our results apply to this form of analysis as well.

## 2. Covariance Structure Analysis

If  $S$  is a symmetric matrix,  $\text{vech}(S)$  is a column vector containing the diagonal and upper diagonal elements  $s_{ij}$  of  $S$  listed in lexicographical order. All but the last line of the following theorem is a well-known result in covariance structure analysis.

**Theorem 1.** *If*

- $x_1, \dots, x_n$  is a sample from a distribution  $\mathcal{D}$  with finite fourth moments, mean  $\mu$ , and covariance matrix  $\Sigma$ ,
- $S = \frac{1}{n} \sum (x_i - \bar{x})(x_i - \bar{x})'$
- $s = \text{vech}(S)$  and  $\sigma = \text{vech}(\Sigma)$ ,

then

$$\sqrt{n}(s - \sigma) \xrightarrow{\mathcal{D}} N(0, \Gamma),$$

where  $\Gamma = \text{cov}(\text{vech}((x - \mu)(x - \mu)'))$  and  $x$  is a sample of size one from  $\mathcal{D}$ .

To discuss the nonsingularity of  $\Gamma$  one must know how  $\Gamma$ , which is a population parameter, is related to the distribution sampled. This is provided by the last line of [Theorem 1](#). We have been unable to find a reference for [Theorem 1](#) that includes this last line. Because of this, a proof is given in the [Appendix](#).

As noted, it is generally assumed that  $\Gamma$  is nonsingular. We will show first that this is always the case when the distribution sampled is nonsingular and, in particular, when the distribution sampled is defined by a density function.

## 3. When the Distribution Sampled Is Nonsingular

We begin with a result about the zeros of a polynomial in  $m$  variables.

**Lemma 1.** *If  $p(x)$  is a polynomial of degree  $d \geq 1$  in  $m$  variables, then the zeros of  $p(x)$  have Lebesgue measure zero.*

*Proof:* The proof is by induction. If  $d = 1$ ,  $p(x) = \ell'x + c$  and  $\ell \neq 0$ . Because  $\ell \neq 0$  the zeros of  $p(x)$  have Lebesgue measure zero.

Assume the theorem holds for  $d = k$  and let  $p(x)$  be a  $k + 1$  degree polynomial in  $m$  variables.

Note that the zeros of  $\dot{p}(x)$  are a closed set. Hence, the  $x$  such that  $\dot{p}(x) \neq 0$  is an open set. It follows from Theorem 5-1 of Spivak (1965) that the zeros of  $p(x)$  for which  $\dot{p}(x) \neq 0$  are a smooth manifold of dimension  $m - 1$ .

It follows from Spivak (1965, problem 5.8) that a manifold of dimension less than  $m$  has Lebesgue measure zero. Thus, the zeros of  $p(x)$  for which  $\dot{p}(x) \neq 0$  have Lebesgue measure zero.

Now consider the zeros of  $p(x)$  for which  $\dot{p}(x) = 0$ . Since  $\dot{p}(x)$  is a  $k$  degree polynomial in  $m$  variables it follows from the induction hypotheses that the zeros of  $\dot{p}(x)$  have Lebesgue measure zero. Thus, the zeros of  $p(x)$  for which  $\dot{p}(x) = 0$  have Lebesgue measure zero. Since the zeros of  $p(x)$  are a union of two sets with Lebesgue measure zero, the zeros of  $p(x)$  have Lebesgue measure zero.  $\square$

A distribution  $\mathcal{D}$  is said to be singular with respect to Lebesgue measure if there is a set with Lebesgue measure zero and probability measure one.

**Theorem 2.** *If the matrix  $\Gamma$  in Theorem 1 is singular, then the distribution  $\mathcal{D}$  in Theorem 1 is singular.*

*Proof:* Since  $s$  and  $\sigma$  do not depend on  $\mu$  we may assume without loss of generality that  $\mu = 0$ . Then  $\Gamma = \text{cov}(\text{vech}(xx'))$ .

Because  $\Gamma$  is singular there is a vector  $\ell \neq 0$  such that  $\ell'\Gamma\ell = 0$ . It follows that  $\text{var}(\ell'\text{vech}(xx')) = \ell'\Gamma\ell = 0$ , and hence there is a  $c$  such that  $\ell'\text{vech}(xx') = c$  with probability one. Let

$$p(x) = \ell'\text{vech}(x'x) - c.$$

Then  $p(x) = 0$  with probability one. Since the components of  $\ell$  are not all zero,  $p(x)$  is a polynomial of degree one or more. It follows from Lemma 1 that the zeros of  $p(x)$  have Lebesgue measure zero. Since they also have probability one, the distribution  $\mathcal{D}$  is singular.  $\square$

The following theorem is an immediate consequence of Theorem 2.

**Theorem 3.** *If the distribution  $\mathcal{D}$  in Theorem 1 is nonsingular, then the matrix  $\Gamma$  in Theorem 1 is nonsingular.*

Thus, as promised, we have shown that  $\Gamma$  is nonsingular whenever the sampling is from a nonsingular distribution. In particular,  $\Gamma$  is nonsingular whenever the sampling is from a distribution defined by a density function.

## 4. When the Distribution Sampled Is Discrete

A natural question is what happens when  $\mathcal{D}$  is singular. From the point of view of applications, the primary singular distributions are the discrete distributions. These are considered here.

**Theorem 4.** *If the distribution  $\mathcal{D}$  in Theorem 1 is discrete with mass points  $d_1, \dots, d_q$ , then  $\Gamma$  is nonsingular if and only if*

$$A = \begin{pmatrix} \text{vech}((d_1 - \mu)(d_1 - \mu)') & \dots & \text{vech}((d_q - \mu)(d_q - \mu)') \\ 1 & \dots & 1 \end{pmatrix}$$

has full row rank.

*Proof:* Let  $e_i = d_i - \mu$ . Then

$$A = \begin{pmatrix} \text{vech}(e_1 e_1') & \dots & \text{vech}(e_q e_q') \\ 1 & \dots & 1 \end{pmatrix}.$$

Let  $e = x - \mu$ ; then

$$\Gamma = \text{cov}(\text{vech}(e e')).$$

Assume  $\Gamma$  is singular. Then there is an  $\ell \neq 0$  such that  $\ell' \Gamma \ell = 0$  and

$$\text{var}(\ell' \text{vech}(e e')) = \ell' \Gamma \ell = 0.$$

This implies

$$\ell' \text{vech}(e e') = c$$

for some  $c$  with probability one. Because  $\mathcal{D}$  is discrete, for all  $i = 1, \dots, q$

$$\ell' \text{vech}(e_i e_i') = c.$$

This can be written in the form

$$(\ell', -c) \begin{pmatrix} \text{vech}(e_1 e_1') & \dots & \text{vech}(e_q e_q') \\ 1 & \dots & 1 \end{pmatrix} = 0,$$

or in terms of  $A$  as

$$(\ell', -c)A = 0.$$

Since  $\ell \neq 0$ , this implies  $A$  does not have full row rank. Thus,  $\Gamma$  singular implies  $A$  does not have full row rank.

Assume now that  $A$  does not have full row rank. Then there is a vector  $a$  and scalar  $b$  such that  $a$  and  $b$  are not both zero and

$$(a', b) \begin{pmatrix} \text{vech}(e_1 e_1') & \dots & \text{vech}(e_q e_q') \\ 1 & \dots & 1 \end{pmatrix} = 0.$$

Note that  $a$  cannot be zero because this would imply  $b = 0$ .

Taking the  $i$ th column of both sides gives

$$a' \text{vech}(e_i e_i') + b = 0$$

for all  $e_i$ . It follows that

$$a' \text{vech}(e e') + b = 0$$

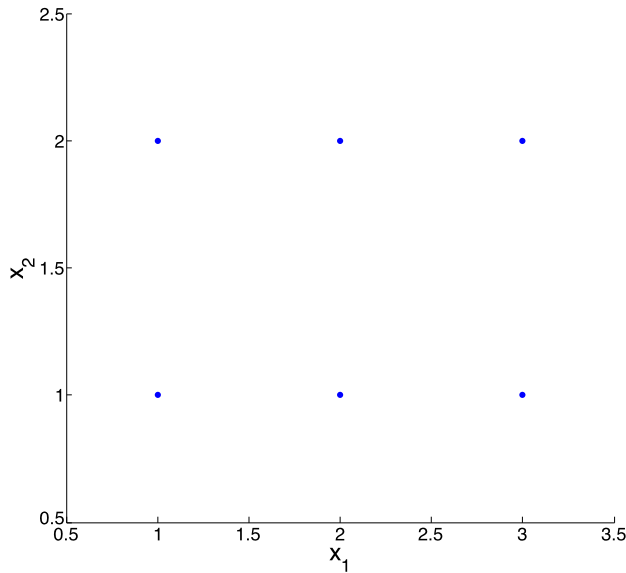


FIGURE 1.  
Mass points.

with probability one. Hence,

$$\text{var}(a' \text{vech}(ee')) = 0.$$

Thus,  $a' \Gamma a = a' (\text{vech}(ee')) a = \text{var}(a' \text{vech}(ee')) = 0$ . Since  $a \neq 0$ ,  $\Gamma$  is singular.

Thus,  $\Gamma$  is singular if and only if  $A$  does not have full row rank; or  $\Gamma$  is nonsingular if and only if  $A$  has full row rank.  $\square$

We will use Theorem 4 to show, as promised, that  $\Gamma$  can be singular even when  $\Sigma$  is nonsingular.

Assume  $\mathcal{D}$  is discrete and has the mass points displayed in Figure 1.

Assume these carry equal probability mass. Clearly,  $\Sigma$  is nonsingular.

Let  $d_1, \dots, d_6$  be the points displayed, then

$$\mu = (d_1 + \dots + d_6)/6 = \begin{pmatrix} 2 \\ 1.5 \end{pmatrix}.$$

This together with the formula for the matrix  $A$  in Theorem 4 gives

$$A = \begin{pmatrix} 1.00 & 1.00 & 0.00 & 0.00 & 1.00 & 1.00 \\ 0.50 & -0.50 & 0.00 & 0.00 & -0.50 & 0.50 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \end{pmatrix}.$$

The singular values of  $A$  are

$$\lambda = \begin{pmatrix} 3.08 \\ 1.00 \\ 0.95 \\ 0.00 \end{pmatrix}.$$

Hence,  $A$  does not have full row rank. It follows from Theorem 4 that  $\Gamma$  is singular.

We have assumed in this example that in the population sampled the probabilities of the design points are exactly equal. In practice, this seems very unlikely. If one assigns random probabilities to the design points the resulting  $A$  matrix always seems to have full row rank. More precisely, when we did this using 1,000 randomly generated probability assignments, all 1,000  $A$  matrices had full row rank, and hence by Theorem 4 had nonsingular  $\Gamma$ . This does not show that  $\Gamma$  is nonsingular with probability one, but the probability must at least be very high when probability masses are assigned at random.

In general, it is difficult to use Theorem 4 to show  $\Gamma$  is nonsingular. One would have to know  $\mu$ , which generally is unknown because the population probability masses are unknown.

As in the previous example, however, one can assign random probability masses to one's mass points to investigate the probability that the distribution sampled in one's specific application has a nonsingular  $\Gamma$ . If a large number of probability assignments all produce a nonsingular  $\Gamma$ , this strongly suggests the population  $\Gamma$  for the distribution under investigation is nonsingular with high probability. This should make one much more comfortable about making this assumption.

A somewhat anecdotal argument for non-singularity of  $\Gamma$  is that it took the authors a long time to find any set of mass points and a probability assignment that would produce a singular  $\Gamma$ . If the mass points are a 3 by 3 array rather than a 2 by 3 array, for example, uniform probabilities give a nonsingular  $\Gamma$ .

## 5. Discussion

We have shown the asymptotic covariance matrix  $\Gamma$  used in covariance structural analysis and structural equation modeling of nonnormal data is generally nonsingular. This is important because this assumption is used to obtain standard errors and goodness of fit tests in standard statistical software.

To date there have been no conditions identified that guarantee or even motivate this nonsingularity assumption in the nonparametric context. Theorem 3 shows that, when sampling from a non-singular distribution,  $\Gamma$  must be nonsingular. The nonsingularity of  $\Gamma$  in the discrete case depends on the probabilities of the mass points in the population sampled. When these are known, Theorem 4 can be used to determine the nonsingularity of  $\Gamma$ . In general, however, they are not known. When this is the case, Theorem 4 can be used to investigate the likelihood that  $\Gamma$  is nonsingular. Our results make one much more comfortable about assuming  $\Gamma$  is nonsingular.

We conjecture that in the discrete case  $\Gamma$  is nonsingular with probability one whenever probability masses are assigned randomly. Our numerical example strongly suggests this is true, but at present we have no proof.

## Appendix: Proof of Theorem 1 and Singular $\Gamma$ Example

### A.1. Proof of Theorem 1

Let  $e_i = x_i - \mu$ . Then  $e_i - \bar{e} = x_i - \bar{x}$  and

$$S_n = \frac{1}{n} \sum (e_i - \bar{e})(e_i - \bar{e})' = \frac{1}{n} \sum e_i e_i' - \bar{e} \bar{e}'$$

and

$$\sqrt{n} S_n = \frac{1}{\sqrt{n}} \sum e_i e_i' - \sqrt{n} \bar{e} \bar{e}'.$$

Since  $\bar{e} = \bar{x} - \mu \xrightarrow{P} 0$  and  $\sqrt{n}\bar{e} = \sqrt{n}(\bar{x} - \mu) \xrightarrow{D} N(0, \Sigma)$ ,  $\sqrt{n}\bar{e}\bar{e}' \xrightarrow{P} 0$ . Thus,

$$\sqrt{n}S_n \stackrel{a}{=} \frac{1}{\sqrt{n}} \sum e_i e_i'$$

and

$$\sqrt{n}(s_n - \sigma) \stackrel{a}{=} \frac{1}{\sqrt{n}} \sum (\text{vech}(e_i e_i') - \sigma). \quad (\text{A.1})$$

Note that the  $\text{vech}(e_i e_i')$  are independent and identically distributed,  $\sigma$  is the common expected value of the  $\text{vech}(e_i e_i') = \text{vech}((x_i - \mu)(x_i - \mu)')$ , and  $\Gamma$  is the common covariance matrix of the  $\text{vech}(e_i e_i')$ . It follows from the central limit theorem that the right-hand side of (A.1) converges in distribution to  $N(0, \Gamma)$ , and hence that

$$\sqrt{n}(s_n - \sigma) \xrightarrow{D} N(0, \Gamma).$$

Moreover,

$$\Gamma = \text{cov}(\text{vech}(e_i e_i')) = \text{cov}(\text{vech}((x_i - \mu)(x_i - \mu)')) = \text{cov}(\text{vech}((x - \mu)(x - \mu)')).$$

### A.2. What Happens When $\Gamma$ Is Singular?

Using a very simple example, we will show that Browne's (1984) extensively used goodness of fit test for CSA can fail completely when  $\Gamma$  is singular. Consider random sampling from the array in Figure 1. We have shown that the asymptotic covariance matrix  $\Gamma$  for the sample covariances  $s$  generated in this way is singular. Consider a covariance structure

$$\gamma(\theta) = \begin{pmatrix} 2/3 \\ \theta \\ 1/4 \end{pmatrix}.$$

When  $\theta = 0$  this is equal to  $\sigma$ . Using a least squares estimate  $\hat{\theta}$  of  $\theta$ , Browne's statistic for testing the goodness of fit of  $\gamma(\theta)$  is

$$T = n(s - \gamma(\hat{\theta}))' U (U' \hat{\Gamma} U)^{-1} U' (s - \gamma(\hat{\theta})),$$

where  $U$  is an orthogonal complement of

$$\Delta = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

the Jacobian of  $\gamma(\theta)$ , and  $\hat{\Gamma}$  is a consistent estimator for  $\Gamma$ . We have used the estimator  $\hat{\Gamma}$  given by Satorra and Bentler (1990, Formula 2.4).

If  $\Gamma$  were nonsingular,  $T$  would have an asymptotic  $\chi^2$  distribution with two degrees of freedom. But in this example  $\Gamma$  is singular. To investigate the performance of Browne's test in this situation we generated  $N = 1000$  samples of size  $n = 100$ , and for each a value of  $T$  was computed. Figure 2 is a Q-Q plot of the values of  $T$  on the corresponding quantiles of the  $\chi_2^2$ . Clearly, the distribution of  $T$  differs greatly from  $\chi_2^2$ . Moreover, in 78 of the 1,000 trials  $T$  could not even be computed because  $\hat{\Gamma}$  was too nearly singular. On these trials, the computer output correctly suggests that  $\Gamma$  is singular. On 92 % of the trials, however, no such warning is produced.

One wonders what would happen in this example if  $\Gamma$  were nonsingular. Assume the array in Figure 1 were replaced by a 3 by 3 array with  $x$  and  $y$  values equal to 1, 2, 3. If these carry equal probability mass,  $\Gamma$  is nonsingular and the Q-Q plot becomes that displayed in Figure 3.

Clearly,  $T$  now is very nearly  $\chi_2^2$  distributed. Moreover, there were no problems in computing the values of  $T$ .

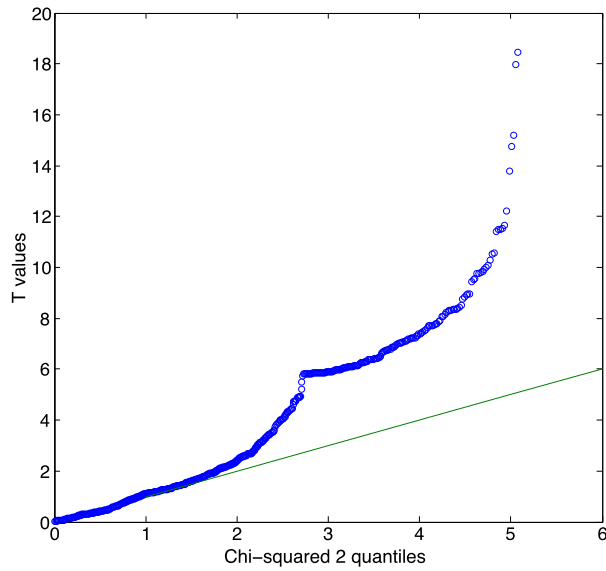


FIGURE 2.  
Q-Q plot when  $\Gamma$  is singular.

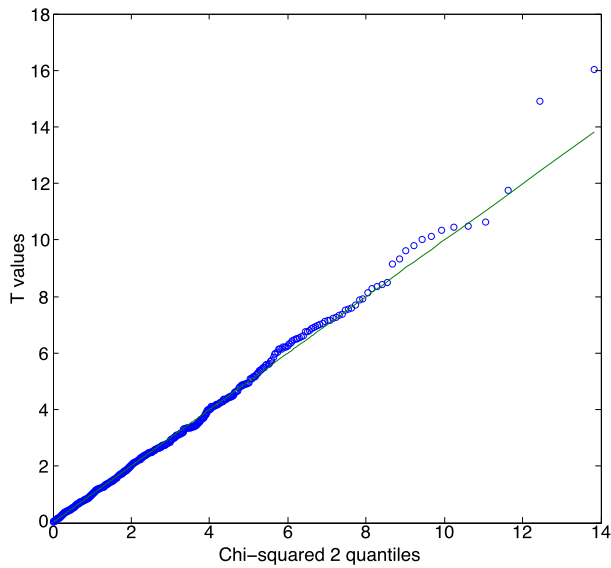


FIGURE 3.  
Q-Q plot when  $\Gamma$  is nonsingular.

#### References

- Bentler, P.M. (1983). Some contributions to efficient statistics in structural models: specification and estimation of structural models. *Psychometrika*, 48, 493–517.
- Browne, M.B. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical & Statistical Psychology*, 37, 62–83.
- Satorra, A. (1989). Alternative test criteria in covariance structure analysis: a unified approach. *Psychometrika*, 54, 131–151.



- Satorra, A., & Bentler, P.M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics & Data Analysis*, *10*, 235–249.
- Spivak, M. (1965). *Calculus on manifolds*. New York: Perseus Book Publishing, LLC
- Yuan, K., & Bentler, P.M. (2001). Effect of outliers on estimators and tests in covariance structure analysis. *British Journal of Mathematical & Statistical Psychology*, *54*, 161–175.

*Manuscript Received: 18 JUL 2012*

*Final Version Received: 14 FEB 2013*

*Published Online Date: 12 SEP 2013*