On Confidence Regions of Structural Equation Models

Jian-Qing Shi*

Department of Computing Science, University of Glasgow, UK

Sik-Yum Lee

Department of Statistics, The Chinese University of Hong Kong

Bo-Cheng Wei

Department of Applied Mathematics, Southeast University, China

*_Address for correspondence_: Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK. Email: shi@dcs.gla.ac.uk

SUMMARY

In this paper, methods in constructing confidence regions for parameters in structural equation models are investigated. Based on the generalized least squares theory, a geometric approach in getting the general region is established. The linear and quadratic approximations of the general case are derived. Moreover, regions for subsets of parameters, and constrained models are obtained. It is shown that the quadratic approximation is better for situations with nonlinear models or nonlinear constraints. Two examples are presented to illustrate the theory.

Key words: Confidence regions; Curvature; Generalized least squares; Nonlinear constraints; Nonlinear models; Structural equation models; Three-mode model.

1 Introduction

Structural equation modelling (Joïeskog, 1978; Bentler, 1983) is an important method in assessing multivariate data in behavioral, medical and social sciences. Under certain mild regularity conditions, the basic asymptotic results for statistical inference of the model have been developed. For example, (Browne, 1974, 1984) established the generalized least squares (GLS) theory of the model and showed that the GLS estimator is asymptotically equivalent to the maximum likelihood (ML) estimator; and Lee & Bentler (1980) extended the normal GLS theory to models with constraints.

Obviously, one of the most important results for practical applications is that the

GLS estimator is asymptotically normal. Based on this result, various null hypotheses about the parameters can be tested. However, it is almost always the case that confidence intervals or regions are "uniformly more informative" than hypothesis tests for making decisions. In fact, confidence intervals (regions) with information on their positions and sizes in the parameter space offered the decision makers the opportunity to consider "practical" significance in addition to "statistical significance". Moreover, confidence intervals can be used to test statistical hypothesis, whereas the results of a hypothesis test cannot answer the questions posed in interval or region estimation. See Burdick & Graybill (1992), and Wonnacott (1987) for additional arguments for preferring confidence intervals or regions to hypothesis tests.

Of course, in analysis of structural equation models, confidence regions of parameters can be constructed based on the asymptotic normality of the GLS or the ML estimator. However, as it will be seen later in Section 2, this kind of confidence region is only a linear approximation of the general one. For some situations with nonlinear constraints or nonlinear models, the bias of this linear approximation is significant (see the discussion in Section 5 via simulation results). In this paper, to improve the linear approximation, a geometric approach in constructing the more general and accurate confidence regions is investigated in the context of the GLS estimation.

In Section 2, the basic idea in constructing the confidence regions based on the GLS theory will be discussed from a geometric point of view. This kind of confidence region will be called the GLS confidence region. The linear and quadratic approximations of the

GLS confidence region are derived. Section 3 extends the results to parameter subsets. The corresponding confidence regions for models with constraints are briefly discussed in Section 4. Illustrative numerical examples are given in Section 5.

The following notations will be used in this article: Suppose $\boldsymbol{A}$ is a p by p symmetric matrix, $\mathrm{vec}(\boldsymbol{A})$ represents the $p^2 \times 1$ column vector formed by stacking the columns of $\boldsymbol{A}$, and $\mathrm{vecs}(\boldsymbol{A})$ represents the column vector obtained from the $p^* = p(p+1)/2$ nonduplicated lower triangular elements of $\boldsymbol{A}$. Let $\boldsymbol{K}_p$ be the $p^2 \times p^*$ transition matrix such that $\mathrm{vecs}(\boldsymbol{A}) = \boldsymbol{K}_p{}'\mathrm{vec}(\boldsymbol{A})$ and $\mathrm{vec}(\boldsymbol{A}) = \boldsymbol{K}_p^{-}{}'\mathrm{vecs}(\boldsymbol{A})$, where $\boldsymbol{K}_p^{-} = (\boldsymbol{K}_p{}'\boldsymbol{K}_p)^{-1}\boldsymbol{K}_p{}'$ (Browne, 1974). The right Kronecker product of matrices is denoted by $\otimes$.

2 Basic Theory of GLS Confidence Region

2.1 Geometric framework of the structural equation model

Suppose that the p by 1 random vectors $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n, \boldsymbol{X}_{n+1}$ are independent and identically distributed as multivariate normal with mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0 = \Sigma(\boldsymbol{\theta}_0)$, where $\boldsymbol{\Sigma}_0$ is a matrix function of a $q \times 1$ vector of unknown parameters $\boldsymbol{\theta}_0$. Let $\boldsymbol{\theta}$ be a vector of mathematical variables that are defined in an open subset $\boldsymbol{\Theta}$ in the q-dimensional Euclidean space R^q, and can take value $\boldsymbol{\theta}_0$ in $\boldsymbol{\Theta}$. Let $\boldsymbol{S}$ be the sample covariance matrix, $\boldsymbol{s} = \mathrm{vecs}(\boldsymbol{S})$, and $\boldsymbol{\sigma}(\boldsymbol{\theta}) = \mathrm{vecs}(\Sigma(\boldsymbol{\theta}))$. The first two derivatives of $\boldsymbol{\sigma}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ are denoted by $\boldsymbol{\Delta}$ and $\boldsymbol{\nabla}$, which are $p^* \times q$ matrix and $p^* \times q \times q$ array, respectively. The following mild regularity conditions are assumed:

(a) $\boldsymbol{\theta}$ is identified; that is $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$.

(b) $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is positive definite in a neighborhood Θ_0 of $\boldsymbol{\theta}_0$.

(c) All the first three derivatives of $\boldsymbol{\sigma}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ are bounded and continuous in Θ_0 and $\boldsymbol{\Delta}(\boldsymbol{\theta})$ is of full column rank in Θ_0.

These regularity conditions are satisfied for the most commonly used structural equation models.

The asymptotic covariance matrix of $\boldsymbol{s}$ is given by

$$\mathrm{Var}(\boldsymbol{s}) = n^{-1}\boldsymbol{g}(\boldsymbol{\theta}) = 2n^{-1}\boldsymbol{K}_p{}'(\boldsymbol{\Sigma}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}(\boldsymbol{\theta}))\boldsymbol{K}_p,$$

where $g^{-1}(\boldsymbol{\theta})$ is the Fisher information matrix of the $\boldsymbol{X}$'s for $\boldsymbol{\sigma}(\boldsymbol{\theta})$. Consider the following structural equation model:

$$\boldsymbol{s} = \boldsymbol{\sigma}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{\varepsilon}$ is the residual vector such that $\boldsymbol{e} = \sqrt{n}\boldsymbol{\varepsilon}$ has the appropriate asymptotically normal distribution at $\boldsymbol{\theta}_0$. (1) is one kind of nonlinear regression model, therefore, some geometric approaches (see e.g., Bates & Watts, 1980, Wei, 1997) can be used. The GLS function (see, Browne, 1974) is given by

$$G(\boldsymbol{\theta}) = 2^{-1}\mathrm{tr}\{(\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))\boldsymbol{V}\}^2 = [\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]'\boldsymbol{W}[\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})], \tag{2}$$

where $\boldsymbol{W} = 2^{-1}\boldsymbol{K}_p^-(\boldsymbol{V} \otimes \boldsymbol{V})\boldsymbol{K}_p^{-\prime}$, and $\boldsymbol{V}$ is a positive definite matrix which converges in probability to $\boldsymbol{\Sigma}_0^{-1}$. Hence, $\boldsymbol{W}$ will converge in probability to $\boldsymbol{g}^{-1}(\boldsymbol{\theta}_0)$. The GLS estimate which minimizes $G(\boldsymbol{\theta})$ is denoted by $\hat{\boldsymbol{\theta}}$. Under the regularity assumptions, it follows from

Browne (1974) that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0},\ 2[\boldsymbol{\Delta}_0{}'\boldsymbol{K}_p^-(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\boldsymbol{K}_p^-{}'\boldsymbol{\Delta}_0]^{-1}), \tag{3}$$

where $\xrightarrow{L}$ denotes convergent in distribution.

Similar to Bates & Watts (1980), Wei (1994) and Wang & Lee (1995), we now introduce a geometric framework for the nonlinear regression model (1) as follows. Take $\boldsymbol{\sigma}$ as coordinates in the Euclidean space R^{p^*}, the solution locus π is defined as $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta})$, a q-dimensional surface in R^{p^*}. The tangent space $T_{\boldsymbol{\theta}}$ of π at $\boldsymbol{\theta}$ is spanned by columns of $\boldsymbol{\Delta}(\boldsymbol{\theta})$. For any two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in R^{p^*}, we define the weighted inner product as $< \boldsymbol{a}, \boldsymbol{b} >= \boldsymbol{a}'\boldsymbol{W}\boldsymbol{b}$, and the corresponding normal space is denoted by $T^*_{\boldsymbol{\theta}}$ (for the geometric concepts such as tangent space and normal space, the reder is referred to a standard textbook of differential geometry, e.g. Millman and Parker, 1977 or certain statistical books, e.g. Seber and Wild, 1989, Appendix B). Suppose that the decomposition of $\boldsymbol{\Delta}(\boldsymbol{\theta})$ is given by

$$\boldsymbol{\Delta}(\boldsymbol{\theta}) = (\boldsymbol{Q}, \boldsymbol{N}) \begin{pmatrix} \boldsymbol{R} \\ \mathbf{0} \end{pmatrix} = \boldsymbol{Q}\boldsymbol{R}, \tag{4}$$

where $\boldsymbol{R}$ and $\boldsymbol{L} = \boldsymbol{R}^{-1}$ are $q \times q$ nonsingular upper triangular matrices, and the columns of $\boldsymbol{Q}$ and $\boldsymbol{N}$ are orthonormal bases for $T_{\boldsymbol{\theta}}$ and $T^*_{\boldsymbol{\theta}}$, respectively. The matrices $\boldsymbol{Q}$ and $\boldsymbol{N}$ satisfy

$$\boldsymbol{Q}'\boldsymbol{W}\boldsymbol{Q} = \boldsymbol{I}_q, \quad \boldsymbol{Q}'\boldsymbol{W}\boldsymbol{N} = \mathbf{0} \ \text{ and } \ \boldsymbol{N}'\boldsymbol{W}\boldsymbol{N} = \boldsymbol{I}_{p^*-q},$$

where $\boldsymbol{I}_q$ and $\boldsymbol{I}_{p^*-q}$ are identity matrices of order q and $p^* - q$, respectively. Based on the reasonings as given in Bates & Watts (1980), and Wang & Lee (1995), the intrinsic curvature array $\boldsymbol{A}^I$ and the parameter-effects curvature array $\boldsymbol{A}^P$ are given as

$$\boldsymbol{A}^I = [\boldsymbol{N}'\boldsymbol{W}][\boldsymbol{U}], \quad \boldsymbol{A}^P = [\boldsymbol{Q}'\boldsymbol{W}][\boldsymbol{U}]; \quad \boldsymbol{U} = \boldsymbol{L}'\boldsymbol{\nabla}\boldsymbol{L}, \tag{5}$$

where $[\cdot][\cdot]$ indicates the array multiplication as defined in Bates & Watts (1980), and Seber & Wild (1989, p.691). It is well-known that $\boldsymbol{A}^I$ is invariant with respect to parameterization and $\boldsymbol{A}^P$ depends on the parameterizations (Bates & Watts, 1980). Finally, based on the results given in Browne (1974) and Wang & Lee (1995), the first two derivatives $\dot{\boldsymbol{G}}(\boldsymbol{\theta})$ and $\ddot{\boldsymbol{G}}(\boldsymbol{\theta})$ of the GLS function $G(\boldsymbol{\theta})$ can be respectively expressed as

$$\dot{\boldsymbol{G}}(\boldsymbol{\theta}) = -2(\sqrt{n})^{-1}\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{e}, \quad \ddot{\boldsymbol{G}}(\boldsymbol{\theta}) = 2\boldsymbol{\Delta}'\boldsymbol{W}\boldsymbol{\Delta} - 2(\sqrt{n})^{-1}[\boldsymbol{e}'\boldsymbol{W}][\boldsymbol{\nabla}]. \tag{6}$$

2.2 GLS confidence region

Consider the following test statistic for the null hypothesis $H_0 : \ \boldsymbol{\theta} = \boldsymbol{\theta}_0$:

$$SD(\boldsymbol{\theta}) = n\{G(\boldsymbol{\theta}) - G(\hat{\boldsymbol{\theta}})\}. \tag{7}$$

As shown in Lee & Bentler (1980, Proposition 5), $SD(\boldsymbol{\theta}_0)$ is asymptoticly chi-squared with degrees of freedom q. Under the normality assumption, the GLS estimate and the ML estimate are asymptotically equivalent, hence it can be shown that $SD(\boldsymbol{\theta})$ is asymptotically equivalent to the likelihood ratio statistic. The confidence region is obtained by inverting the corresponding test. Specifically, based on the above $SD(\boldsymbol{\theta})$, the $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$ is given by

$$C(\boldsymbol{S}) = \{\boldsymbol{\theta} : SD(\boldsymbol{\theta}) \leq \chi^2(q, \alpha)\}, \tag{8}$$

where $\chi^2(q,\alpha)$ is the upper α percentage point of chi-square distribution with degrees of freedom q. This is called the GLS confidence region for $\boldsymbol{\theta}$. Since the deviation $SD(\boldsymbol{\theta})$ strongly depends on the nonlinear functions $\sigma_{ij}(\boldsymbol{\theta})$, it is usually complicated in the parameter space and is difficult to evaluate. Thus, as pointed out by Hamilton, Watts & Bates (1982), it is valuable to obtain adequate approximate confidence regions.

2.3 Linear approximation of the GLS confidence region

Since

$$\boldsymbol{\sigma}(\boldsymbol{\theta}) \approx \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}) + \boldsymbol{\Delta}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \quad \text{and} \quad \dot{\boldsymbol{G}}(\hat{\boldsymbol{\theta}}) = \mathbf{0},$$

let $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\hat{\boldsymbol{\theta}})$, we have

$$\begin{aligned}
& G(\boldsymbol{\theta}) \\
\approx\ & [\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\theta}) - \boldsymbol{\Delta}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})]'\boldsymbol{W}[\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\theta}) - \boldsymbol{\Delta}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})] \\
=\ & [\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\theta})]'\boldsymbol{W}[\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\theta})] + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\boldsymbol{\Delta}'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) - 2(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\boldsymbol{\Delta}'\boldsymbol{W}(\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\theta})) \\
=\ & G(\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\boldsymbol{\Delta}'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).
\end{aligned}$$

Therefore, the linear approximation of GLS confidence region (8) is

$$\{\boldsymbol{\theta}:\ n(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\boldsymbol{\Delta}'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \le \chi^2(q,\alpha)\}. \tag{9}$$

This approximate region is equivalent to the region based on the large-sample normality of $\hat{\boldsymbol{\theta}}$ in (3). In the case of very large samples, this is a very efficient approach to calculate confidence region. However, in the case of medium or small samples, it should be noted that this linear approximation may be far away from the GLS confidence region if the nonlinearity of the model is strong. Combined with examples, more discussion will be given in Section 5.

2.4 Quadratic approximation of the GLS confidence region

To improve the above linear approximation, a quadratic approximation of the general confidence region will be derived by means of a geometric approach in this subsection. The idea is to consider a nonlinear transformation $\boldsymbol{\tau}$ from $\boldsymbol{\Theta}$ to a new parameter space $\boldsymbol{\Gamma}$, such that $\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta})$. For any subspace $C(\boldsymbol{S}) \subset \boldsymbol{\Theta}$ such that $P_{\boldsymbol{\theta}}\{\boldsymbol{\theta} : \boldsymbol{\theta} \in C(\boldsymbol{S})\} \geq 1 - \alpha$, $\boldsymbol{\tau}$ tranforms $C(\boldsymbol{S})$ to $K(\boldsymbol{S}) \subset \boldsymbol{\Gamma}$ and that

$$P_{\boldsymbol{\theta}}\{\boldsymbol{\theta} : \boldsymbol{\tau}(\boldsymbol{\theta}) \in K(\boldsymbol{S})\} = P_{\boldsymbol{\theta}}\{\boldsymbol{\theta} : \boldsymbol{\theta} \in C(\boldsymbol{S})\} \geq 1 - \alpha,$$

where $K(\boldsymbol{S}) = \tau(C(\boldsymbol{S}))$. Based on the reasonings given in Hamilton, Watts & Bates (1982), and Wei (1994), this nonlinear transformation is taken as

$$\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta}) = \sqrt{n}\boldsymbol{Q}'\boldsymbol{W}\{\boldsymbol{\sigma}(\boldsymbol{\theta}) - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})\}. \tag{10}$$

In the sequel, all quantities $\boldsymbol{Q}, g, \boldsymbol{R}, \boldsymbol{L}$, etc. are all evaluated at $\hat{\boldsymbol{\theta}}$. As a new parameter vector, $\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta})$ represents an one-to-one mapping from the parameter space $\boldsymbol{\Theta}$ to the tangent space $T_{\hat{\boldsymbol{\theta}}}$ and $\boldsymbol{Q}\boldsymbol{\tau}(\boldsymbol{\theta})$ is just the projection of $\sqrt{n}\{\boldsymbol{\sigma}(\boldsymbol{\theta}) - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})\}$ onto the tangent space. The coordinates in $\boldsymbol{\tau}$ provide a natural reference system for the solution locus $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta})$. We may construct confidence regions for parameters $\boldsymbol{\theta}$ in terms of the coordinates $\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta})$ by using quadratic approximations. Now let the inverse of $\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta})$ be

$\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\tau})$, then we have $\boldsymbol{\tau}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ and $\boldsymbol{\theta}(\mathbf{0}) = \hat{\boldsymbol{\theta}}$.

Lemma 1. The derivatives $\ddot{\boldsymbol{G}}(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}$ and the first two derivatives of $\boldsymbol{\theta}(\boldsymbol{\tau})$ with respect to $\boldsymbol{\tau}$ at $\boldsymbol{\tau} = \mathbf{0}$ are respectively given by

$$\ddot{\boldsymbol{G}}(\hat{\boldsymbol{\theta}}) \overset{a}{=} 2\boldsymbol{R}'(\boldsymbol{I}_q - \boldsymbol{B}/\sqrt{n})\boldsymbol{R}, \tag{11}$$

$$\left(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\tau}'}\right)_{\boldsymbol{\tau}=\mathbf{0}} = \boldsymbol{L}/\sqrt{n}, \quad \text{and} \quad \left(\frac{\partial^2 \boldsymbol{\theta}}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'}\right)_{\boldsymbol{\tau}=\mathbf{0}} = -n^{-1}[\boldsymbol{L}][\boldsymbol{A}^P], \tag{12}$$

where $\overset{a}{=}$ stands for "is asymptotically equivalent to", and $\boldsymbol{B} = [\boldsymbol{e}'\boldsymbol{W}\boldsymbol{N}][\boldsymbol{A}^I]$ evaluated at $\hat{\boldsymbol{\theta}}$.

The proof of Lemma 1 is given in the Appendix. The following theorem, which gives the main result for the GLS confidence region, is a consequence of the lemma.

Theorem 1. Under the regularity conditions, the geometric approximation of $100(1-\alpha)\%$ GLS region of $\boldsymbol{\theta}$ is expressed as

$$K(\boldsymbol{S}) = \{\boldsymbol{\theta} : \boldsymbol{\tau}'(\boldsymbol{I}_q - \boldsymbol{B}/\sqrt{n})\boldsymbol{\tau} \leq \chi^2(q, \alpha), \quad \boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta})\} \tag{13}$$

Proof. From Lemma 1, the first two derivatives of $SD(\boldsymbol{\theta}(\boldsymbol{\tau}))$ with respect to $\boldsymbol{\tau}$ are given by

$$\left\{\frac{\partial SD(\boldsymbol{\theta}(\boldsymbol{\tau}))}{\partial \boldsymbol{\tau}}\right\}_{\boldsymbol{\tau}=\mathbf{0}} = \left\{n(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\tau}'})'\dot{\boldsymbol{G}}(\boldsymbol{\theta})\right\}_{\hat{\boldsymbol{\theta}}} = \mathbf{0};$$

and

$$\left\{\frac{\partial^2 SD(\boldsymbol{\theta}(\boldsymbol{\tau}))}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'}\right\}_{\boldsymbol{\tau}=\mathbf{0}} = \left\{n(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\tau}'})'\ddot{\boldsymbol{G}}(\boldsymbol{\theta})(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\tau}'})\right\}_{\hat{\boldsymbol{\theta}}} = 2(\boldsymbol{I}_q - \boldsymbol{B}/\sqrt{n}).$$

Applying the second order Taylor series expansion to (8), we have

$$SD(\boldsymbol{\theta}) \approx \boldsymbol{\tau}'(\boldsymbol{\theta})(\boldsymbol{I}_q - \boldsymbol{B}/\sqrt{n})\boldsymbol{\tau}(\boldsymbol{\theta}), \tag{14}$$

and hence (13) is valid. $\square$

Under the regularity conditions, $\ddot{\boldsymbol{G}}(\hat{\boldsymbol{\theta}})$ is positive definite, and so is $\boldsymbol{I}_q - \boldsymbol{B}/\sqrt{n}$. Therefore, the approximate GLS region given in (13) is an illipsoid in tangent space $T_{\hat{\boldsymbol{\theta}}}$. Moreover, Since $\boldsymbol{B}$ depends only on the intrinsic curvature array $\boldsymbol{A}^I$, $K(\boldsymbol{S})$ is invariant to parameterization in the tangent space. Expression (13) is similar to the results of Hamilton, Watts & Bates (1982), and Wei(1994) in the context of ML estimation of the exponential family model.

From the above theorem, it is not difficult to construct the confidence region of $\boldsymbol{\tau}$ in the tangent plane because it is expressed as an ellipsoid as given in (13). However, for statistical inference and interpretation in specific applications, confidence region of $\boldsymbol{\theta}$ is more valuable. Unfortunately, it is generally difficult to calculate the inverse transformation which maps the tangent plane ellipsoid into the parameter space, and hence difficult to obtain the confidence region of $\boldsymbol{\theta}$ in the original parameter space. In practice, as suggested by Bates & Watts (1981), the following expression that based on the quadratic Taylor series is commonly used:

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + \boldsymbol{L}(\boldsymbol{\tau}/\sqrt{n} - \frac{1}{2n}\boldsymbol{\tau}'\boldsymbol{A}^P\boldsymbol{\tau}). \tag{15}$$

Using this approximation of the inverse transformation, the bound of confidence region in terms of $\boldsymbol{\tau}$ can be transferred to the bound of confidence region for $\boldsymbol{\theta}$. Since the confidence region in tangent space is an illipsoid, it is easy to obtain this confidence region. As the result, confidence region for $\boldsymbol{\theta}$ can also be constructed efficiently. Examples given in Section 5 show that this approximation works quite well and fast. We recommend to use this approach in practice, especially in the case of medium or samll samples.

3 Regions for Subsets of Parameters

Under some practical situations, the joint confidence region of $\boldsymbol{\theta}$ with high dimension may be difficult to construct and interpret. Hence, in this section, we consider confidence regions for subsets of the parameters. Let $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1{}', \boldsymbol{\theta}_2{}')$ be a partition of $\boldsymbol{\theta}$, where $\boldsymbol{\theta}_1$ is $q_1 \times 1$, while $\boldsymbol{\theta}_2$ is the $q_2 \times 1$ parameter vector of interest, and $q = q_1 + q_2$. Moreover, let $\boldsymbol{\tau}' = (\boldsymbol{\tau}_1{}', \boldsymbol{\tau}_2{}')$, $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$, $\boldsymbol{R} = (\boldsymbol{R}_{ij})$ and $\boldsymbol{B} = (\boldsymbol{B}_{ij})$ for $i, j = 1, 2$ be the corresponding partitions that conform to the partitioning of $\boldsymbol{\theta}$. Let $\tilde{\boldsymbol{\theta}}' = (\tilde{\boldsymbol{\theta}}_1'(\boldsymbol{\theta}_2),\ \boldsymbol{\theta}_2{}')$, where $\tilde{\boldsymbol{\theta}}_1(\boldsymbol{\theta}_2)$ minimized $G(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ for a fixed $\boldsymbol{\theta}_2$. Then the analogous statistic

$$SD(\tilde{\boldsymbol{\theta}}) = n\{G(\tilde{\boldsymbol{\theta}}) - G(\hat{\boldsymbol{\theta}})\} \tag{16}$$

is asymptotically chi-squared with degrees of freedom q_2 under the null hypothesis H_0 : $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_{02}$; see Lee & Bentler (1980). The $100(1-\alpha)\%$ GLS confidence region of $\boldsymbol{\theta}_2$ is expressed as

$$C_2(\boldsymbol{S}) = \{\boldsymbol{\theta}_2 : SD(\tilde{\boldsymbol{\theta}}) \leq \chi^2(q_2, \alpha)\}. \tag{17}$$

Since

$$\frac{\partial \boldsymbol{\sigma}(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}_2} = \boldsymbol{\Delta}_1 \frac{\partial \tilde{\boldsymbol{\theta}}_1}{\partial \boldsymbol{\theta}_2} + \boldsymbol{\Delta}_2, \quad \text{and} \quad \frac{\partial \tilde{\boldsymbol{\theta}}_1}{\partial \boldsymbol{\theta}_2} = -(\boldsymbol{\Delta}_1 \boldsymbol{W} \boldsymbol{\Delta}_1)^{-1} \boldsymbol{\Delta}_1 \boldsymbol{W} \boldsymbol{\Delta}_2,$$

the linear approximating region can be derived through the following Taylor's expansion:

$$\begin{aligned}
& \tilde{\boldsymbol{\sigma}}(\boldsymbol{\theta}_2) = \boldsymbol{\sigma}(\tilde{\boldsymbol{\theta}}) \\
\approx \;& \tilde{\boldsymbol{\sigma}}(\hat{\boldsymbol{\theta}}_2) + \frac{\partial \boldsymbol{\sigma}(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}_2}(\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2) \\
= \;& \tilde{\boldsymbol{\sigma}}(\hat{\boldsymbol{\theta}}_2) + \boldsymbol{W}^{-1/2}(I - \boldsymbol{W}^{1/2}(\boldsymbol{\Delta}_1 \boldsymbol{W} \boldsymbol{\Delta}_1)^{-1} \boldsymbol{\Delta}_1 \boldsymbol{W}^{1/2}) \boldsymbol{W}^{1/2} \boldsymbol{\Delta}_2 (\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2) \\
= \;& \tilde{\boldsymbol{\sigma}}(\hat{\boldsymbol{\theta}}_2) + \boldsymbol{W}^{-1/2} \boldsymbol{Q}_1 \boldsymbol{W}^{1/2} \boldsymbol{\Delta}_2 (\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2),
\end{aligned}$$

where $\boldsymbol{Q}_1 = I - \boldsymbol{W}^{1/2}(\boldsymbol{\Delta}_1 \boldsymbol{W} \boldsymbol{\Delta}_1)^{-1} \boldsymbol{\Delta}_1 \boldsymbol{W}^{1/2}$ is a projection matrix. Therefore, by the similar derivations of (9), the linear approximation region is expressed as

$$\{\boldsymbol{\theta}_2 : \; n(\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2)' \boldsymbol{\Delta}_2' \boldsymbol{W}^{1/2} \boldsymbol{Q}_1 \boldsymbol{W}^{1/2} \boldsymbol{\Delta}_2 (\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2) \le \chi^2(q_2, \alpha)\}. \tag{18}$$

To obtain the more general GLS confidence region, we need to get the approximate tangent space projection of $C_2(\boldsymbol{S})$. Note that equations (10) and (14) are still true at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, hence,

$$\tilde{\boldsymbol{\tau}} = \boldsymbol{\tau}(\tilde{\boldsymbol{\theta}}) = \sqrt{n} \boldsymbol{Q}' \boldsymbol{W} \{\boldsymbol{\sigma}(\tilde{\boldsymbol{\theta}}) - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})\}, \tag{19}$$

$$SD(\tilde{\boldsymbol{\theta}}) \approx \boldsymbol{\tau}'(\tilde{\boldsymbol{\theta}})(\boldsymbol{I}_q - \boldsymbol{B}/\sqrt{n}) \boldsymbol{\tau}(\tilde{\boldsymbol{\theta}}). \tag{20}$$

From these results, the following theorem is true:

Theorem 2. Under the regularity conditions, the geometric approximation of $100(1-$

α)% GLS region for $\boldsymbol{\theta}_2$ is expressed as

$$K_2(\boldsymbol{S}) = \{\boldsymbol{\theta}_2 : \tilde{\boldsymbol{\tau}}_2'(\boldsymbol{I}_2 - \boldsymbol{T}/\sqrt{n})\tilde{\boldsymbol{\tau}}_2 \leq \chi^2(q_2, \alpha), \quad \tilde{\boldsymbol{\tau}}_2 = \boldsymbol{\tau}_2(\boldsymbol{\theta}_2)\}, \tag{21}$$

where $\boldsymbol{T} = \boldsymbol{B}_{22} + \boldsymbol{B}_{21}(\sqrt{n}\boldsymbol{I}_1 - \boldsymbol{B}_{11})^{-1}\boldsymbol{B}_{12}$, $\boldsymbol{I}_1 = \boldsymbol{I}_{q_1}$, and $\boldsymbol{I}_2 = \boldsymbol{I}_{q_2}$.

Proof. From the approximations to (10) and (19), we have

$$\boldsymbol{\tau} \approx \boldsymbol{Q}'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = \sqrt{n}\boldsymbol{R}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

and

$$\tilde{\boldsymbol{\tau}} \approx \sqrt{n}\boldsymbol{R}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}});$$

respectively. $\boldsymbol{R}$ is upper triangular, $\boldsymbol{R}_{21} = 0$, and

$$\tilde{\boldsymbol{\tau}}_2 = \boldsymbol{\tau}_2 \approx \sqrt{n}\boldsymbol{R}_{22}(\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2).$$

On the other hand, it follows from (11) that

$$\dot{\boldsymbol{G}}(\tilde{\boldsymbol{\theta}}) \approx \ddot{\boldsymbol{G}}(\hat{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = 2\boldsymbol{R}'(\boldsymbol{I}_q - \boldsymbol{B}/\sqrt{n})\tilde{\boldsymbol{\tau}}/\sqrt{n}.$$

Since $(\partial G/\partial\boldsymbol{\theta}_1)_{\tilde{\boldsymbol{\theta}}} = \mathbf{0}$, we get $(\boldsymbol{I}_1 - \boldsymbol{B}_{11}/\sqrt{n})\tilde{\boldsymbol{\tau}}_1 - \boldsymbol{B}_{12}\tilde{\boldsymbol{\tau}}_2\sqrt{n} = \mathbf{0}$. Hence,

$$\tilde{\boldsymbol{\tau}}_1 = (\sqrt{n}\boldsymbol{I}_1 - \boldsymbol{B}_{11})^{-1}\boldsymbol{B}_{12}\tilde{\boldsymbol{\tau}}_2, \quad \tilde{\boldsymbol{\tau}}_2 = \boldsymbol{\tau}_2. \tag{22}$$

Substituting these results into (20) yields (21). □

It can be seen that $K_2(\boldsymbol{S})$ is an illipsoid in the tangent space and invariant under reparameterizations. To get the confidence region in the parameter space, the following quadratic Taylor series can be used as a mapping from the tangent plane to the parameter space:

$$\boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_2 + \boldsymbol{L}_{22}(\boldsymbol{\tau}_2/\sqrt{n} - \frac{1}{2n}\boldsymbol{\tau}_2{}'\boldsymbol{K}\boldsymbol{\tau}_2), \tag{23}$$

where $\boldsymbol{K} = [\boldsymbol{Q}_2{}'\boldsymbol{W}][\boldsymbol{L}_{22}{}'\boldsymbol{\nabla}_2\boldsymbol{L}_{22}]$, and $\boldsymbol{\nabla}_2$ is $p^* \times q_2 \times q_2$ portion of $\boldsymbol{\nabla}$ consisting of the second derivatives of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ with respect to the parameters of interest.

Both Hamilton (1986) and Wei (1994) had obtained the confidence regions for subsets of parameters based on the score statistics. For completeness, we give this kind of confidence region for structural equation models. Viewing from (6) that $G(\boldsymbol{\theta})$ can be regarded as a quasi-likelihood function, an analogous score test statistic is given as follows.

Lemma 2. The square derivation $\boldsymbol{S}D(\tilde{\boldsymbol{\theta}})$ given in (16) can be equivalently expressed as

$$SD(\tilde{\boldsymbol{\theta}}) \stackrel{a}{=} \frac{n}{2}\left\{(\frac{\partial G}{\partial \boldsymbol{\theta}_2})'\boldsymbol{J}^{22}(\frac{\partial G}{\partial \boldsymbol{\theta}_2})\right\}_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}, \tag{24}$$

where $\boldsymbol{J}^{22}$ is the lower right corner of $\boldsymbol{J}^{-1}$ and $\boldsymbol{J} = 2\boldsymbol{\Delta}'\boldsymbol{W}\boldsymbol{\Delta}$ is obtained from $E(\ddot{G}(\boldsymbol{\theta}))$.

Theorem 3. Under regularity conditions, the approximate $100(1-\alpha)\%$ score-based region for $\boldsymbol{\theta}_2$ is expressed as

$$K_2'(\boldsymbol{S}) = \{\boldsymbol{\theta}_2 : \tilde{\boldsymbol{\tau}}_2'(\boldsymbol{I}_2 - \boldsymbol{T}/\sqrt{n})^2\tilde{\boldsymbol{\tau}}_2 \leq \chi^2(q_2, \alpha),\ \tilde{\boldsymbol{\tau}}_2 = \boldsymbol{\tau}_2(\boldsymbol{\theta}_2)\}. \tag{25}$$

The proofs of above lemma and theorem are given in the Appendix. Expression (25) is similar to those of Hamilton (1986) and Wei (1994) in the ML approach, but our model assumptions and derivations are quite different from theirs.

4 Regions in Constrained Models

Consider the GLS function $G(\boldsymbol{\theta})$ given in (2) subject to functional constraints $\boldsymbol{h}(\boldsymbol{\theta}) = \boldsymbol{0}$, where $\boldsymbol{h}(\boldsymbol{\theta})$ is a $r \times 1$ vector of differentiable functions in $\boldsymbol{\Theta}_0$, a neighborhood of $\boldsymbol{\theta}_0$. The reparameterization approach (Lee & Bentler, 1980) was applied to get the confidence region of $\boldsymbol{\theta}$ under this situation.

Let $s = q - r$, there exists a $s \times 1$ vector $\boldsymbol{\gamma}$ such that $\boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\gamma})$ and $\boldsymbol{h}(\boldsymbol{g}(\boldsymbol{\gamma})) = \boldsymbol{0}$. Let $\partial \boldsymbol{g}(\boldsymbol{\gamma})/\partial \boldsymbol{\gamma}' = \boldsymbol{K}$, and $\hat{\gamma}$ be the unconstrained GLS estimate of $\boldsymbol{\gamma}$ that minimizes

$$G_c(\boldsymbol{\gamma}) = 2^{-1}\mathrm{tr}\{[\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{g}(\boldsymbol{\gamma}))]\boldsymbol{V}\}^2, \tag{26}$$

and $\hat{\boldsymbol{\theta}}_c = \boldsymbol{g}(\hat{\gamma})$. The solution locus π_c corresponding to the constrained model is $\{\pi_c : \boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta}), \quad \boldsymbol{h}(\boldsymbol{\theta}) = \boldsymbol{0}\}$, which can also be expressed as the following unconstrained form

$$\pi_c : \boldsymbol{\sigma} = \boldsymbol{\sigma}_c(\boldsymbol{\gamma}) = \boldsymbol{\sigma}(\boldsymbol{g}(\boldsymbol{\gamma})), \tag{27}$$

in a s-dimensional surface. Hence, this induces an equivalent unconstrained model, and results of previous sections can be applied.

To introduce similar curvature measures for π_c as in Section 1, let the first two derivatives of $\boldsymbol{\sigma}_c(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ be denoted by $\boldsymbol{\Delta}_c$ and $\boldsymbol{\nabla}_c$, respectively. Moreover, in the sequel, let $\boldsymbol{R}_c, \boldsymbol{L}_c, \boldsymbol{Q}_c, \boldsymbol{N}_c, \boldsymbol{U}_c$ and $\boldsymbol{A}_c^I$ be the corresponding matrices defined in (4) and

(5) in the context of the constrained model. The corresponding intrinsic arrays are

$$\boldsymbol{A}_c^I = [\boldsymbol{N}_c^{'}\boldsymbol{W}][\boldsymbol{U}_c], \quad \boldsymbol{U}_c = \boldsymbol{L}_c^{'}\nabla_c\boldsymbol{L}_c. \tag{28}$$

To compute $\boldsymbol{A}_c^I$, we need to find the $p^* \times (p^* - s)$ matrix $\boldsymbol{N}_c$, whose columns are orthonormal bases of normal space of π_c. It can be seen from (27) that $\boldsymbol{\Delta}_c = \boldsymbol{\Delta K} = \boldsymbol{QRK}$, so the tangent space $\boldsymbol{T}_c$ of π_c is spanned by columns of $\boldsymbol{QRK}$. Moreover, since $\boldsymbol{HK} = \boldsymbol{0}$ (see Lee & Bentler, 1980), we have

$$(\boldsymbol{QL}^{'}\boldsymbol{H}^{'})^{'}\boldsymbol{W}(\boldsymbol{QRK}) = \boldsymbol{HK} = \boldsymbol{0}.$$

Thus, the normal space $\boldsymbol{T}_c^*$ of π_c is spanned by columns of $\boldsymbol{QL}^{'}\boldsymbol{H}^{'}$ and $\boldsymbol{N}$. From these results, we can get $\boldsymbol{N}_c$ and $\boldsymbol{A}_c^I$.

To obtain the confidence regions for constrained models, the transformation (10) can be applied to (26) and (27), which leads to the following one-to-one mapping

$$\boldsymbol{\varphi} = \boldsymbol{\varphi}(\boldsymbol{\gamma}) = \sqrt{n}\boldsymbol{Q}_c^{'}\boldsymbol{W}\{\boldsymbol{\sigma}_c(\boldsymbol{\gamma}) - \boldsymbol{\sigma}_c(\hat{\boldsymbol{\gamma}})\}. \tag{29}$$

Then, applying the results in Theorem 1 to (26) and (27) yields a GLS confidence region

$$\boldsymbol{K}_c(\boldsymbol{S}) = \{\boldsymbol{\gamma} : \boldsymbol{\varphi}^{'}(\boldsymbol{I}_s - \boldsymbol{B}_c/\sqrt{n})\boldsymbol{\varphi} \le \chi^2(s, \alpha),\ \boldsymbol{\varphi} = \boldsymbol{\varphi}(\boldsymbol{\gamma})\}, \tag{30}$$

where $\boldsymbol{B}_c = [\boldsymbol{e}_c{}'\boldsymbol{W}\boldsymbol{N}_c][\boldsymbol{A}_c^I]$ evaluated at $\hat{\boldsymbol{\theta}}_c$, and $\boldsymbol{e}_c = \sqrt{n}[\boldsymbol{s} - \boldsymbol{\sigma}_c(\boldsymbol{\gamma})]$. On the other hand, it follows from (27) that $\boldsymbol{\sigma}_c(\boldsymbol{\gamma}) = \boldsymbol{\sigma}(\boldsymbol{g}(\boldsymbol{\gamma}))$, $\boldsymbol{g}(\hat{\gamma}) = \hat{\boldsymbol{\theta}}_c$ and $\boldsymbol{g}(\boldsymbol{\gamma}) = \boldsymbol{\theta}$, so (29) and (30) become

$$\begin{aligned} \boldsymbol{\varphi}_c &= \boldsymbol{\varphi}_c(\boldsymbol{\theta}) = \sqrt{n}\boldsymbol{Q}_c^{'}\boldsymbol{W}\{\boldsymbol{\sigma}(\boldsymbol{\theta}) - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}_c)\}, \quad \text{and} \\ \boldsymbol{K}_c(\boldsymbol{S}) &= \{\boldsymbol{\theta} : \boldsymbol{\varphi}_c^{'}(\boldsymbol{I}_s - \boldsymbol{B}_c/\sqrt{n})\boldsymbol{\varphi}_c \le \chi^2(s, \alpha),\ \boldsymbol{\varphi}_c = \boldsymbol{\varphi}_c(\boldsymbol{\theta})\}. \end{aligned}$$

This is a GLS region of $\boldsymbol{\theta}$ subject to the constraint $\boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0}$. Similar results in Theorems 2 and 3 can also be extended to the constrained model.

5 Numerical Examples

We will use two examples to demonstrate the accuracy of the linear and quadratic approximation of the GLS confidence region.

5.1 Factor analysis model with nonlinear constraints

In this example, a factor analysis model

$$\boldsymbol{y} = \boldsymbol{\Lambda} \boldsymbol{f} + \boldsymbol{\varepsilon}$$

with following structure is used:

$$\boldsymbol{\Lambda}' = \begin{bmatrix} 1 & 0 & \lambda & 0 \\ 0 & 1 & 0 & 1/(1-\lambda) \end{bmatrix}, \quad cov(\boldsymbol{f}) = \boldsymbol{\Phi} = \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}, \quad cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi},$$

where $\boldsymbol{\Psi}$ is a diagonal matrix. For illustration purpose, elements in $\boldsymbol{\Psi}$ as well as the one's and zero's in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Phi}$ are treated as fixed parameters. Hence, the only unknown parameters are λ and ϕ, and their population values were taken to be $\lambda_0 = 1.5$ and $\phi_0 = 0.6$. Fifty five observations were generated based on the above model from the appropriate multivariate normal distribution with the given populations values. From these observations, the GLS estimates, which were computed via the Newton-Raphson algorithm, are equal to $\hat{\lambda} = 1.51$ and $\hat{\phi} = 0.53$.

The GLS confidence region introduced in Section 2, as well as its quadratic approximation to the geometric confidence region and the linear approximation region were obtained. The GLS confidence region was obtained via the following Monte Carlo method. Values of $\boldsymbol{\theta} = (\lambda, \phi)'$ were uniformly drawn from a selected large enough area, then $SD(\boldsymbol{\theta})$ were calculated from (7). If a $SD(\boldsymbol{\theta})$ value satisfied the condition in (8), the corresponding $\boldsymbol{\theta}$ was included as a point in the confidence region. After a sufficient large number of $\boldsymbol{\theta}$ was checked, the skeleton of confidence region was determined. This confidence region may be regarded as a standard region for comparison. Please note that the computation burden is very heavy and the accuracy can not be guaranteed by this approach. Even though the computer power has been rapidly increased nowadays, it still becomes time consuming for the problem with large dimension of $\boldsymbol{\theta}$.

In obtaining the quadratic approximation of the geometric confidence region, the tangent plane ellipsoid given in (13) of Theorem 1 was computed and then a map of the quadratic Taylor series (15) was used to get the region in the parameter space. The linear approximation confidence region was constructed via (9). Figure 1 gives the 95% confidence regions. Apparently, the linear approximation is bias; while the quadratic approximation geometric confidence region is close to the GLS confidence region.

Figure 1 is about here.

The coverage probabilitis were also calculated via a standard Monte Carlo simulation with 5000 replications. For the GLS confidence region, its quadratic approximation, and its linear approximation, the coverage probablities are 0.912, 0.918 and 0.927 respectively, which are all close to $1 - \alpha = 0.95$. Moreover, we plot all 5000 estimates of simulation in Figure 1. They stand for the empirical distribution of the estimates. It shows that the linear approximation confidence region is skewed to left, while the confidence region by GLS method and the quadratic approximation are more close to the "shape" of the empirical distribution. Since the computational burden of GLS confidence region is heavy, while its quadratic and linear approximations are easy to be achieved, hence, in practice, the quadratic approximation of the GLS is recommended.

5.2 Example based on a three-mode model

The data reported in Bentler & Lee (1979) is used here as a second example to illustrate the theory. This data set consisted of four personality traits measured by three methods of measurement. The sample correlation matrix of this multitrait-multimethod data set was also reanalyzed by Lee(1985), and Bentler, Poon & Lee (1988) via the following 3-mode model:

$$\boldsymbol{\Sigma} = (\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{G}'(\boldsymbol{A} \otimes \boldsymbol{B})' + \boldsymbol{\Psi},$$

with

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I}_A \\ \boldsymbol{a}' \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{I}_B \\ \boldsymbol{b}' \end{bmatrix},$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are 3×2 and 4×3 matrices repectively, $\boldsymbol{I}_A(2 \times 2)$ and $\boldsymbol{I}_B(3 \times 3)$ are fixed identity matrices, $\boldsymbol{a}' = (a_1, a_2)$ and $\boldsymbol{b}' = (b_1, b_2, b_3)$ are parameters, $\boldsymbol{\Phi}(6 \times 6)$ is a fixed identity matrix, and the upper triangular elements of $\boldsymbol{G}(6 \times 5)$ are fixed at zero while the remaining elements of $\boldsymbol{G}$ are free parameters, and $\boldsymbol{\Psi}$ is diagonal with $\boldsymbol{\Psi}(1,1)$ and $\boldsymbol{\Psi}(5,5)$ fixed at zero. For brevity, in this paper, we fixed all the parameters in $\boldsymbol{G}$ and $\boldsymbol{\Psi}$ at the GLS estimates given in Bentler, Poon & Lee (1988).

By the theory discussed in this paper, confidence region for all five parameters in $\boldsymbol{a}$ and $\boldsymbol{b}$ can be obtained. However, in order to display the results clearly, we constructed confidence regions in a 2-dimensional plane for a subset of parameters, say $\boldsymbol{\theta}_2 = (a_1, b_1)'$ via the method developed in Section 3. The GLS confidence region obtained via the Monte Carlo method, the quadratic approximation of the GLS confidence region and the linear approximation of the confidence region are displayed in Figure 2. It shows that the region obtained by linear approximation is an ellipsoid. The region obtained by quadratic approximation, taken into account the nonlinearlity of the model, is more close to the standard GLS confidence region than by linear approximation.

Figure 2 is about here.

6 Discussion

Most existing asymptotic theory in structural equation modelling is derived from the first order of the Taylor's series expansion. Hence, this kind of theory is only the linear approximation of the general case. As a result, to deal with nonlinear constraints or the more complex nonlinear models, better methods that based on the more general theory are necessary. In this article, using a geometric approach in the GLS estimation, we derive some better methods in constructing confidence regions. It is shown that the improvement over the linear approximation is significant for nonlinear models and nonlinear constraints. Analogues theory in the context of ML analysis with the normal assumption or the asymptotically distribution-free analysis with arbitrary distribution (see, e.g. Browne, 1984) can be derived using the similar geometric approach as proposed in this article.

As we have shown in the last section, the linear approximation is the simplest approach to construct a confidence region. If the sample size is sufficiently large, the region obtained by linear approximation is close to the GLS confidence region. Therefore, this approach can be used in the case of large samples. In the case of medium or small samples, we suggest to use quadratic approximation, especially for the model with strong nonlinearity. The reader is refered to Bates and Watts (1980) for measuring the nonlinearity. Monte Carlo method can be used to construct an almost exact GLS confidence region. However, it becomes time consuming if the dimension of $\boldsymbol{\theta}$ is large.

Acknowledgement

Comments from the referee and Dr. Moustaki have led to an improved version. One of us (BCW) is partially supported by NSF of China.

Appendix

A.1 Proof of Lemma 1

It follows from (4), (5) and (6) that $\ddot{\boldsymbol{G}}(\boldsymbol{\theta})$ can be expressed as

$$\begin{aligned}
\ddot{\boldsymbol{G}}(\boldsymbol{\theta}) &\overset{a}{=} 2\boldsymbol{\Delta}'\boldsymbol{W}\boldsymbol{\Delta} - 2(\sqrt{n})^{-1}[\boldsymbol{e}'\boldsymbol{W}][\nabla] \\
&= 2\boldsymbol{R}'\boldsymbol{R} - 2\boldsymbol{R}'[\boldsymbol{e}'\boldsymbol{W}][\boldsymbol{U}]\boldsymbol{R}/\sqrt{n} \\
&= 2\boldsymbol{R}'(\boldsymbol{I}_q - [\boldsymbol{e}'\boldsymbol{W}][\boldsymbol{U}]/\sqrt{n})\boldsymbol{R}. \qquad \text{(A1)}
\end{aligned}$$

It follows from (6) that $\dot{\boldsymbol{G}}(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$ implies $(\boldsymbol{e}'\boldsymbol{W}\boldsymbol{Q})_{\hat{\boldsymbol{\theta}}} \overset{a}{=} \boldsymbol{0}$, so from (5) we have

$$\begin{aligned}
\{[\boldsymbol{e}'\boldsymbol{W}][\boldsymbol{U}]\}_{\hat{\boldsymbol{\theta}}} &= \{[e'\boldsymbol{W}(\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{W} + \boldsymbol{N}\boldsymbol{N}'\boldsymbol{W})][\boldsymbol{U}]\}_{\hat{\boldsymbol{\theta}}} \\
&= \{[\boldsymbol{e}'\boldsymbol{W}\boldsymbol{N}\boldsymbol{N}'\boldsymbol{W}][\boldsymbol{U}]\}_{\hat{\boldsymbol{\theta}}} \\
&= \{[\boldsymbol{e}'\boldsymbol{W}\boldsymbol{N}][\boldsymbol{A}^I]\}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{B}.
\end{aligned}$$

From this and (A1) we get (11). To prove (12), we just note from (10), (4) and (5) that

$$\left(\frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\theta}'}\right)_{\hat{\boldsymbol{\theta}}} = \sqrt{n}\boldsymbol{R}, \qquad \left(\frac{\partial^2\boldsymbol{\tau}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)_{\hat{\boldsymbol{\theta}}} = \sqrt{n}\boldsymbol{R}'\boldsymbol{A}^P\boldsymbol{R}. \qquad \text{(A2)}$$

Then

$$\frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\tau}'} = \frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\theta}'}\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\tau}'} = \boldsymbol{I}_q,$$

and hence

$$\left(\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\tau}'}\right)_{\boldsymbol{\tau}=\mathbf{0}} = \boldsymbol{L}/\sqrt{n}.$$

Moreover,

$$\frac{\partial^2\boldsymbol{\tau}}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'} = \left(\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\tau}'}\right)' \left(\frac{\partial^2\boldsymbol{\tau}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) \left(\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\tau}'}\right) + \left[\frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\theta}'}\right] \left[\frac{\partial^2\boldsymbol{\theta}}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}\right] = \mathbf{0}.$$

From this and (A2), we get (12).

A.2 Proof of Lemma 2

From (16), and the second order Taylor series expansion of $SD(\tilde{\boldsymbol{\theta}})$ at $\tilde{\boldsymbol{\theta}}$, we have

$$SD(\tilde{\boldsymbol{\theta}}) = -n\dot{\boldsymbol{G}}'(\tilde{\boldsymbol{\theta}})\delta(\boldsymbol{\theta}) - \frac{1}{2}n\delta'(\boldsymbol{\theta})\ddot{\boldsymbol{G}}(\tilde{\boldsymbol{\theta}})\delta(\boldsymbol{\theta}) + O_p(n^{-\frac{1}{2}}), \tag{A3}$$

where $\boldsymbol{\delta}(\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}$. To get the partitioned expression of (A3), let $\boldsymbol{\delta}(\boldsymbol{\theta}) = (\boldsymbol{\delta}_1{}'(\boldsymbol{\theta}), \boldsymbol{\delta}_2{}'(\boldsymbol{\theta}))'$, where $\delta_1(\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}_1 - \tilde{\boldsymbol{\theta}}_1$, $\delta_2(\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2$, and write expansion of $\dot{\boldsymbol{G}}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ at $\tilde{\boldsymbol{\theta}}$ as

$$\dot{\boldsymbol{G}}(\tilde{\boldsymbol{\theta}}) + \ddot{\boldsymbol{G}}(\tilde{\boldsymbol{\theta}})\delta(\boldsymbol{\theta}) + O_p(n^{-1}) = \mathbf{0}.$$

It yields

$$\dot{\boldsymbol{G}}(\tilde{\boldsymbol{\theta}}) = -\ddot{\boldsymbol{G}}(\tilde{\boldsymbol{\theta}})\boldsymbol{\delta}(\boldsymbol{\theta}) + O_p(n^{-1}), \tag{A4}$$

$$\boldsymbol{\delta}(\boldsymbol{\theta}) = \{\ddot{\boldsymbol{G}}(\tilde{\boldsymbol{\theta}})\}^{-1}\dot{\boldsymbol{G}}(\tilde{\boldsymbol{\theta}}) + O_p(n^{-1}). \tag{A5}$$

Now let $\dot{\boldsymbol{G}} = (\dot{\boldsymbol{G}}_1', \dot{\boldsymbol{G}}_2')'$, $\ddot{\boldsymbol{G}} = (\boldsymbol{G}_{ij})$, and $\ddot{\boldsymbol{G}}^{-1} = (\boldsymbol{G}^{ij})$ for $i, j = 1, 2$. Since $\dot{\boldsymbol{G}}_1(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$, from (A4), we get

$$\boldsymbol{G}_{11}(\tilde{\boldsymbol{\theta}})\boldsymbol{\delta}_1(\boldsymbol{\theta}) + \boldsymbol{G}_{12}(\tilde{\boldsymbol{\theta}})\boldsymbol{\delta}_2(\boldsymbol{\theta}) = O_p(n^{-1}). \tag{A6}$$

From (A5) and (A6), we get

$$\begin{aligned}\boldsymbol{\delta}_2(\boldsymbol{\theta}) &= -\boldsymbol{G}^{22}(\tilde{\boldsymbol{\theta}})\dot{\boldsymbol{G}}_2(\tilde{\boldsymbol{\theta}}) + O_p(n^{-1}),\\ \boldsymbol{\delta}_1(\boldsymbol{\theta}) &= -\boldsymbol{G}_{11}^{-1}(\tilde{\boldsymbol{\theta}})\boldsymbol{G}_{12}(\tilde{\boldsymbol{\theta}})\boldsymbol{\delta}_2(\boldsymbol{\theta}) + O_p(n^{-1}).\end{aligned}$$

Since $\dot{\boldsymbol{G}}(\boldsymbol{\theta}) = O_p(n^{-\frac{1}{2}})$ (see (6)), $\boldsymbol{\delta}(\boldsymbol{\theta}) = O_p(n^{-\frac{1}{2}})$, substituting the above results into the partitioned expression of (A3) and by a little calculation yield

$$SD(\tilde{\boldsymbol{\theta}}) \overset{a}{=} \frac{1}{2} n \dot{\boldsymbol{G}}_2^{'}(\tilde{\boldsymbol{\theta}})\boldsymbol{G}^{22}(\tilde{\boldsymbol{\theta}})\dot{\boldsymbol{G}}_2(\tilde{\boldsymbol{\theta}}). \tag{A7}$$

On the other hand, it follows from (6) that $\ddot{\boldsymbol{G}}(\boldsymbol{\theta}) \overset{a}{=} 2\boldsymbol{\Delta}^{'}\boldsymbol{W}\boldsymbol{\Delta} = \boldsymbol{J}$ and $\ddot{\boldsymbol{G}}^{-1} \overset{a}{=} \boldsymbol{J}^{-1}$. Hence, $\boldsymbol{G}^{22} \overset{a}{=} \boldsymbol{J}^{22}$, and from (A7), we get (24).

A.3 Proof of Theorem 3

It follows from (6) that

$$\partial \boldsymbol{G}/\partial \boldsymbol{\theta}_2 = -2(\sqrt{n})^{-1}\boldsymbol{\Delta}_2^{'}\boldsymbol{W}\boldsymbol{e} \overset{a}{=} -2(\sqrt{n})^{-1}\boldsymbol{\Delta}_2^{'}\boldsymbol{W}\boldsymbol{e}.$$

Let

$$\boldsymbol{P} = \boldsymbol{\Delta}(\boldsymbol{\Delta}'\boldsymbol{W}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^{'}\boldsymbol{W}, \quad \boldsymbol{P}_1 = \boldsymbol{\Delta}_1(\boldsymbol{\Delta}_1^{'}\boldsymbol{W}\boldsymbol{\Delta}_1)^{-1}\boldsymbol{\Delta}_1^{'}\boldsymbol{W}$$

and $\boldsymbol{P}_1^* = \boldsymbol{I}_{p*} - \boldsymbol{P}_1$, then $\boldsymbol{J} = 2\boldsymbol{\Delta}^{'}\boldsymbol{W}\boldsymbol{\Delta}$ gives

$$\boldsymbol{J}^{22} = 2^{-1}(\boldsymbol{\Delta}_2^{'}\boldsymbol{W}\boldsymbol{P}_1^*\boldsymbol{\Delta}_2)^{-1}.$$

Substituting the above results into (24) yields

$$SD(\tilde{\boldsymbol{\theta}}) \overset{a}{=} \{\boldsymbol{e}^{'}\boldsymbol{W}\boldsymbol{\Delta}_2(\boldsymbol{\Delta}_2^{'}\boldsymbol{W}\boldsymbol{P}_1^*\boldsymbol{\Delta}_2)^{-1}\boldsymbol{\Delta}_2^{'}\boldsymbol{W}\boldsymbol{e}\}_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}.$$

By similar derivations as shown in Hamilton (1986) and Wei (1994), we get

$$SD(\tilde{\boldsymbol{\theta}}) \stackrel{a}{=} \{\boldsymbol{e}'\boldsymbol{W}(\boldsymbol{P}-\boldsymbol{P}_1)\boldsymbol{e}\}_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}. \tag{A8}$$

To get (25) from this, we need to compute $(\boldsymbol{e}'\boldsymbol{W}\ \boldsymbol{P}\boldsymbol{e})_{\tilde{\boldsymbol{\theta}}}$ and $(e'\boldsymbol{W}\boldsymbol{P}_1\boldsymbol{e})_{\tilde{\boldsymbol{\theta}}}$ in terms of the parameter $\tilde{\boldsymbol{\tau}}$ defined in (16) and (19). Expanding $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta}(\boldsymbol{\tau}))$ at $\boldsymbol{\tau} = \mathbf{0}$, it follows from chain rule and Lemma 1 that

$$\boldsymbol{\sigma}(\boldsymbol{\theta}(\boldsymbol{\tau})) \approx \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}) + (\sqrt{n})^{-1}\boldsymbol{Q}\boldsymbol{\tau} + (2n)^{-1}N(\boldsymbol{\tau}'\boldsymbol{A}^I\boldsymbol{\tau}). \tag{A9}$$

Moreover, since $\tilde{\boldsymbol{e}} = \sqrt{n}\{\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}(\tilde{\boldsymbol{\tau}}))\}$, we have

$$\tilde{\boldsymbol{e}} \approx \hat{\boldsymbol{e}} - \boldsymbol{Q}\tilde{\boldsymbol{\tau}} - (2\sqrt{n})^{-1}\boldsymbol{N}(\tilde{\boldsymbol{\tau}}'\boldsymbol{A}^I\tilde{\boldsymbol{\tau}}).$$

Further, it follows from (A9) that

$$\partial\boldsymbol{\sigma}/\partial\boldsymbol{\tau}' \approx (\sqrt{n})^{-1}\boldsymbol{Q} + (n^{-1})\boldsymbol{N}(\boldsymbol{A}^I\boldsymbol{\tau}),$$

which can be used to obtain the projection matrices $\boldsymbol{P}$ and $\boldsymbol{P}_1$ because

$$\boldsymbol{\Delta} = \frac{\partial\boldsymbol{\sigma}}{\partial\boldsymbol{\theta}'} = \frac{\partial\boldsymbol{\sigma}}{\partial\boldsymbol{\tau}'}\cdot\frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\theta}'}.$$

Using all the above results and some calculations, we get

$$SD(\tilde{\boldsymbol{\theta}}) \approx \tilde{\boldsymbol{\tau}}'(\boldsymbol{I}_q - \boldsymbol{B}/\sqrt{n})(\boldsymbol{I}_q - \boldsymbol{E}_1)(\boldsymbol{I}_q - \boldsymbol{B}/\sqrt{n})\tilde{\boldsymbol{\tau}},$$

where $\boldsymbol{E}_1$ is a $q \times q$ matrix with $\boldsymbol{I}_{q1}$ in the upper left corner and zeros elsewhere. Substituting (22) into the above expression yields (25).

REFERENCES

Bates, D.M. and Watts, D.G. (1980). Relative curvature measures of nonlinearity (with discussion). *J. Roy. Statist. Soc. Ser. B. 40*, 1-25.

Bates, D.M. and Watts, D.G. (1981). Parameter transformation for improved approximate confidence regions in nonlinear least squares. *Annals of Statistics, 9*, 1152-1167.

Bentler, P.M. (1983). Some contributions to efficient statistics for structure models: specification and estimation of moment structures. *Psychometrika, 48*, 493-517.

Bentler, P.M. and Lee, S.Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology, 32*, 87-104.

Bentler, P.M., Poon, W.Y. and Lee, S.Y. (1988). Generalized multimode latent variable models: Implementation by standard programs. *Computational Statistics & Data Analysis, 6*, 107-118.

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 in the analysis of covariance structure. *British Journal of Mathematical and Statistical Psychology, 37*, 62-83.

Burdick, R.K. & Graybill, F.A. (1992) *Confidence Intervals on Variance Components.* New York: Marcel Dekker, Inc.

Hamilton, D.C. (1986). Confidence regions for parameter subsets in nonlinear regression. *Biometrika, 73*, 57-64.

Hamilton, D.C., Watts, D.G., and Bates, D.M. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. *Annals of Statistics, 10*, 386-393.

Jöreskog, K.G. (1978). Structure analysis of covariance and correlation matrices. *Psychometrika, 43*, 443-477.

Lee, S.Y. (1985). On testing functional constraints in structural equation models. *Biometrika, 72*, 125-131.

Lee, S.Y. and Bentler, P.M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Statistical Journal, 14*, 121-136.

Millman, R.S. and Parker, G.D. (1977). *Elements of Differential Geometry.* Englewood Cliffs, NJ: Prentice Hall.

Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression.* Wiley, New York.

Wang, S.J. and Lee, S.Y. (1995). A geometric approach of the generalized least-squares estimation in analysis of covariance structures. *Statistical & Probability Letters, 24*, 39-47.

Wei, B.C. (1994). On confidence regions of embedded models in regular parametric families (A geometric approach). *Australian Journal of Statistics, 36*, 327-328.

Wei, B.C. (1997). *Exponential Family Nonlinear Models.* Springer, Singapore.

Wonnacott,T. (1987). Confidence intervals or hypothesis tests? *Journal of Applied Statistics, 14*, 195-201.

figure captions:

Figure 1. Confidence regions of λ and ϕ, where '.' represents 5000 GLS estimates of (λ, ϕ) of simulation, '—' represents the 'standard' GLS region, '$\cdots$ ' represents the quadratic approximation of geometric confidence region, and '- $\cdot$ - ' represents the linear approximation.

Figure 2. Confidence regions of a_1 and b_1 for Personality data, where '—' represents the 'standard' GLS region, '$\cdot \cdot \cdot$ ' represents the quadratic approximation of geometric confidence region, '- $\cdot$ - ' represents the linear approximation, and '+' represents the GLS estimate of (a_1, b_1).

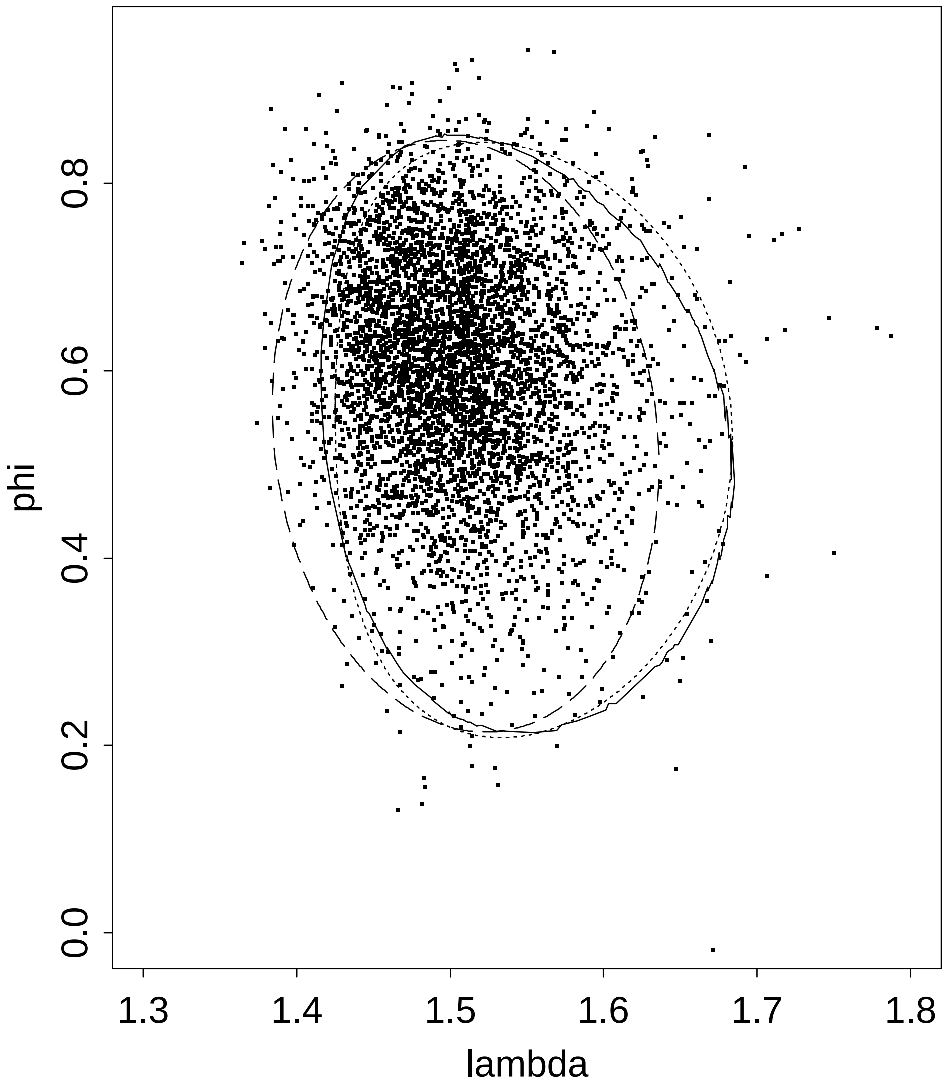

Figure 1

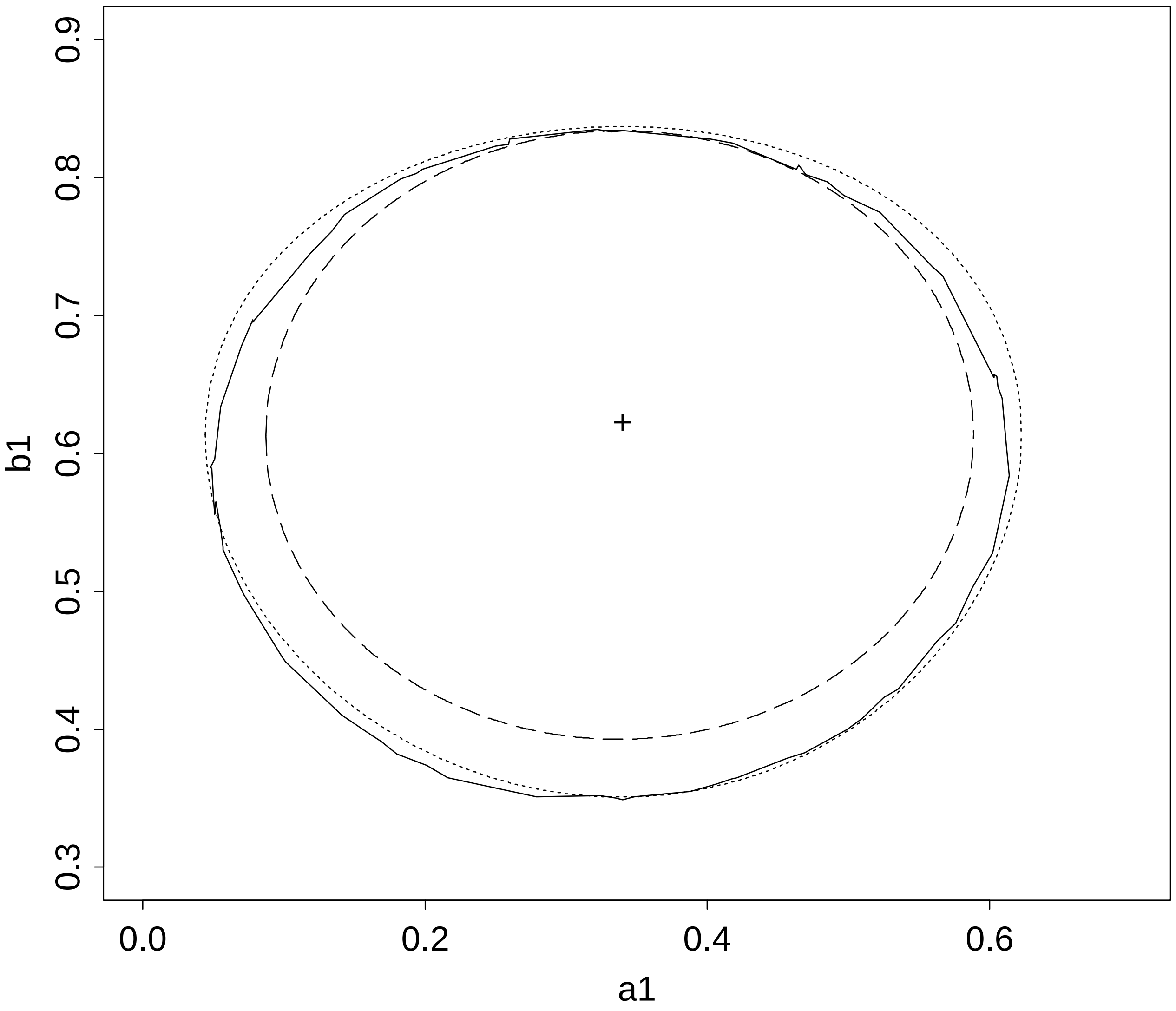

Figure 2

.

Jian Qing Shi received the B.Sc. (Nanjing University, China) in mathematics, M.Sc (Southeast University, China) and Ph.D (The Chinese University of Hong Kong) both in statistics. The research interests include *latent variable models and covariance structural analysis, meta-analysis, influence analysis*, and *nonlinear system control* (see the details in web site: www.dcs.gla.ac.uk/∼shi).

Sik Yum Lee is a chair professor of the statistics department, The Chinese University of Hong Kong. His research interests are in *structural equation modeling, local influence analysis*, and *nonparametric estimation.*

Bo-Cheng Wei is a Professor of Statistics in Southeast University, Nanjing, People's Republic of China. He is RSS fellow in UK and the members of ASA and IMS in USA. Wei's research interests are *nonlinear models* and *diagnostics methods.*