Statistica Sinica **11**(2001), 135-145

ASYMPTOTIC THEORY OF TWO-LEVEL STRUCTURAL EQUATION MODELS WITH CONSTRAINTS

Wenyang Zhang and Sik-Yum Lee

The London School of Economics and Political Science
and The Chinese University of Hong Kong

Abstract: In the context of a general two-level structural equation model with an unbalanced design and small samples at the individual levels, maximum likelihood theory is developed for estimation of the unknown parameters subject to functional constraints. It is shown that the constrained maximum likelihood estimates are consistent and asymptotically normal. A goodness-of-fit statistic is established to test the validity of the constraints. The asymptotic results are illustrated with an example.

Key words and phrases: Asymptotic distribution, goodness-of-fit statistic, maximum likelihood, two-level structural equation models.

1. Introduction

Structural equation modelling (Jöreskog (1978), Bentler (1983)) is a multivariate technique for studying covariances and other relationships among observed and latent variables. Based on the assumption that observed data are independent, it has been applied widely in behavioral, medical and social sciences; see, for example, Cuttance and Ecob (1987), and citations in the well-known packages LISREL 8 (Jöreskog and Sörbom (1996)) and EQS (Bentler (1992)). However in practice, data commonly have a hierarchical structure: students are nested within schools, workers within factories, patients within hospitals, etc. Because individuals within a group are expected to share certain influencing factors and to produce correlated observations, the assumption of independence among observations is violated. In recent years, to take into account the correlated structure of the data, theoretical and computational results of some multilevel models have been established. Theoretically, the maximum likelihood (ML) theory for various special cases with balanced designs and invariance within group structures was developed by McDonald and Goldstein (1989), and Muthèn (1989). Statistical properties of the generalized least squares (GLS) and ML estimation for models with unbalanced designs were established by Lee (1990). Computationally, for some specific models, Lee and Poon (1992) showed that the "multi-sample" option of LISREL 8 or EQS can be used to obtain a consistent estimate which is less efficient than the GLS or the ML estimate. Longford and Muthèn (1992) derived a scoring-type algorithm to obtain the ML estimator in a special case. Recently, Lee and Poon (1998) formulated the ML estimation of the general model as an incomplete-data problem by treating the random vectors at the group level as hypothetical missing data, and a procedure based on the EM algorithm was developed.

On the other hand, the use of equality constraints on the unknown parameters plays an important role even in the analysis of single level structural equation models. Appropriate constraints help to identify the model, give more freedom in defining the covariance structures and provide more meaningful interpretation. See, for example, Lee (1985). LISREL 8 and EQS have options to estimate parameters subject to linear constraints. In the analysis of multilevel models that involve between-group and different within-group structures, the importance of constraints is more significant because relationships of parameters across and among the between-group and various within-group structures can be assessed. However, statistical theory for multilevel models with constraints is still not established. The asymptotic properties developed in Lee (1990) are based on the assumption that the sample size in each group is sufficiently large. In some practical applications, due to the nature of the study or the sampling scheme, available observations in some of the groups may be small.

Our main objective is to develop constrained ML theory for general two-level structural equation models with unbalanced designs. Asymptotic results are established based on a large number of groups, but the sample sizes in the groups can be large or small. Our development will be concentrated on two-level models, but it can be generalized to higher-level models via similar reasoning.

The paper is organized as follows. The main theory will be developed in Section 2, an example is given in Section 3, and a discussion is given in Section 4. Proofs are in the Appendix.

The following notation will be used. If $A = (a_{ij})$ is an $m \times m$ symmetric matrix, $Vec(A)$ is the $m^2 \times 1$ column vector formed by stacking the columns of A, $\lambda_{min}(A)$ represents the minimal eigenvalue of A, and $\|A\| = (\sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij}^2)^{\frac{1}{2}}$. For symmetric matrices A_1 and A_2, $A_1 > A_2$ if and only if $A_1 - A_2$ is positive definite. An identity matrix of order m is denoted by I_m. For simplicity, we denote constants by "a", they may differ at different places.

2. Constrained ML Estimation of Two-level Structural Equation Model

Consider the following general two-level model:

$$x_{gi} = u_g + v_{gi}, \quad g = 1, \ldots, G, \quad i = 1, \ldots, N_g, \tag{2.1}$$

where x_{gi} is a $p \times 1$ observed random vector, u_g is a latent random vector at the group level, and v_{gi} is a latent random vector varying at the individual level. In this model, the random vectors u_g, $g = 1, \ldots, G$, are independently and identically distributed (i.i.d.) as $N[0, \Omega_B(\theta_0)]$; for each g, the random vectors v_{gi}, $i = 1, \ldots, N_g$, are i.i.d. as $N[0, \Omega_{gw}(\theta_0)]$ and independent of u_g, where θ_0 is a $q \times 1$ vector of true parameters. Based on the assumption that G and the N_g are sufficiently large, asymptotic properties of the ML and the GLS estimators of θ_0 have been established by Lee (1990). In this article, we first derive the asymptotic properties of the unconstrained ML estimator without assumptions on the N_g. Based on these preliminary results, we study the constrained ML estimation with the following constraints:

$$h(\theta_0) = (h_1(\theta_0), \ldots, h_r(\theta_0)) = 0, \tag{2.2}$$

where $h_1, \ldots, h_r$ are differentiable functions.

Let θ be the vector of unknown parameters associated with θ_0. The following regularity conditions about the covariance models and the constraints are assumed.

(c1) Let $\Delta_{gw}(\theta) = \partial\Omega_{gw}(\theta)/\partial\theta$, $\Delta_B(\theta) = \partial\Omega_B(\theta)/\partial\theta$, $\Omega_g(\theta) = \Omega_{gw}(\theta) + N_g\Omega_B(\theta)$, $\Delta_g(\theta) = \partial\Omega_g(\theta)/\partial\theta$, and

$$H_G(\theta) = \sum_{g=1}^{G} \Big\{(N_g - 1)\Delta_{gw}(\theta)\Big[\Omega_{gw}^{-1}(\theta) \otimes \Omega_{gw}^{-1}(\theta)\Big]\Delta_{gw}(\theta)' + \Delta_g(\theta)\Big[\Omega_g^{-1}(\theta) \otimes \Omega_g^{-1}(\theta)\Big]\Delta_g(\theta)'\Big\}.$$

There exist matrices $\Sigma(\theta)$ and Σ_{**} such that $G^{-1}H_G(\theta) \to \Sigma(\theta) > \Sigma_{**} > 0$ as $G \to \infty$, in some neighbourhood of θ_0. (The matrix $H_G(\theta)$ can be interpreted as the Fisher information matrix (see Lee (1990)).

(c2) There exists a constant M such that

$$\|\frac{\partial^2\Delta_B(\theta)}{\partial\theta\partial\theta'}\| \le M, \quad \|\Delta_{gw}(\theta)\| \le M, \quad \|\frac{\partial\Delta_{gw}(\theta)}{\partial\theta}\| \le M,$$
$$\text{and} \quad \|\frac{\partial^2\Delta_{gw}(\theta)}{\partial\theta\partial\theta'}\| \le M$$

in some neighbourhood of θ_0.

(c3) All elements in the matrix $\partial h(\theta)/\partial\theta$ are continuous in an open neighbourhood of θ_0.

(c4) The rank of the matrix $\partial h(\theta)/\partial\theta|_{\theta=\theta_0}$ is r.

Let $Z_g = (x'_{g1}, \ldots, x'_{gN_g})'$. From the definition of the model, maximizing the log-likelihood function based on $\{Z_g,\ g = 1, \ldots, G\}$ is equivalent to minimizing the discrepancy function: $F(\theta) = \sum_{g=1}^{G}\{(N_g - 1)\log|\Omega_{gw}(\theta)| + (N_g - 1)tr[\Omega_{gw}^{-1}(\theta)S_g] + \log|\Omega_g(\theta)| + N_g tr[\Omega_g^{-1}(\theta)\bar{x}_g\bar{x}'_g]\}$, where $\bar{x}_g$ and S_g are the sample mean and the sample covariance matrix of the gth group, respectively. Hence, the constrained ML estimate $\tilde{\theta}$ of θ_0 is defined as the vector that minimizes $F(\theta)$ subject to $h(\theta) = 0$. Differentiate $F(\theta)$ with respect to θ twice to find

$$\begin{aligned}\dot{F}(\theta) = &-\sum_{g=1}^{G}\{(N_g - 1)\Delta_{gw}(\theta)[\Omega_{gw}^{-1}(\theta)\otimes\Omega_{gw}^{-1}(\theta)]Vec[S_g - \Omega_{gw}(\theta)]\\ &+\Delta_g(\theta)[\Omega_g^{-1}(\theta)\otimes\Omega_g^{-1}(\theta)]Vec[N_g\bar{x}_g\bar{x}'_g - \Omega_g(\theta)]\},\\ \ddot{F}(\theta) = &-\sum_{g=1}^{G}\{(N_g - 1)A_g(\theta)[I_q\otimes Vec(S_g - \Omega_{gw}(\theta))]\\ &+B_g(\theta)[I_q\otimes Vec(N_g\bar{x}_g\bar{x}'_g - \Omega_g(\theta))]\} + H_G(\theta),\end{aligned}$$

where $A_g(\theta) = \partial[\Delta_{gw}(\theta)(\Omega_{gw}^{-1}(\theta)\otimes\Omega_{gw}^{-1}(\theta))]/\partial\theta$, $B_g(\theta) = \partial[\Delta_g(\theta)(\Omega_g^{-1}(\theta)\otimes\Omega_g^{-1}(\theta))]/\partial\theta$.

For simplicity $\Delta_{gw} = \Delta_{gw}(\theta_0)$, $\Omega_{gw} = \Omega_{gw}(\theta_0)$, $\Omega_g = \Omega_g(\theta_0)$, $\Delta_g = \Delta_g(\theta_0)$, and $H_G = H_G(\theta_0)$. When N_g is large, the positive definite matrix H_G is a good approximation to the Hessian matrix $\ddot{F}(\theta_0)$ because $S_g - \Omega_{gw}$ and $N_g\bar{x}_g\bar{x}'_g - \Omega_g$ are small. We now consider a less straightforward situation without assumptions on the N_g. Let $D_G(\delta) = \{\theta : \|H_G^{\frac{1}{2}}(\theta - \theta_0)\| \le \delta\}$, and $\partial D_G(\delta) = \{\theta : \|H_G^{\frac{1}{2}}(\theta - \theta_0)\| = \delta\}$.

Lemma 1. *For any* $\delta > 0$, $\sup_{\theta\in D_G(\delta)}\|H_G^{-\frac{1}{2}}\ddot{F}(\theta)H_G^{-\frac{1}{2}} - I_q\| \xrightarrow{P} 0$, *where* $H_G^{-\frac{1}{2}}$ *is the inverse square root of* H_G.

Corollary 1. *For all* $\delta > 0$, $P\{\ddot{F}(\theta) - 2^{-1}H_G \ge 0$, *for all* $\theta\in D_G(\delta)\} \longrightarrow 1$.

Corollary 2. *If* $\{\theta_G,\quad G = 1, 2, \ldots\}$ *is any sequence of random vectors such that* $\|H_G^{\frac{1}{2}}(\theta_G - \theta_0)\| = O_P(1)$, *we have* $\|H_G^{-\frac{1}{2}}\ddot{F}(\theta_G)H_G^{-\frac{1}{2}} - I_q\| \xrightarrow{P} 0$.

Let $\hat{\theta}$ be the ML estimate of θ_0 without constraints. The consistency of $\hat{\theta}$ is a consequence of the following lemma, and also useful in deriving our results on constrained estimation.

Lemma 2. $\|H_G^{\frac{1}{2}}(\hat{\theta} - \theta_0)\| = O_P(1)$.

Results on the asymptotic distribution of $\hat{\theta}$ are essential to constructing confidence intervals and to testing various hypotheses concerning θ_0. In contrast to Lee (1990), the asymptotic normality of $\hat{\theta}$ will be derived below without any assumption on the N_g.

Lemma 3. *The asymptotic distribution of* $H_G^{-\frac{1}{2}}\dot{F}(\theta_0)$ *is* $N(0, 2I_q)$.

Theorem 1. *The asymptotic distribution of* $H_G^{\frac{1}{2}}(\theta_0)(\hat{\theta} - \theta_0)$ *is* $N(0, 2I_q)$.

Proofs for the lemmas and corollaries are omitted, they are available from the authors. The proofs for the main theorems are given in the Appendix. Lemmas 2 and 3, and Theorem 1 on the unconstrained ML estimation, are not given in Lee (1990).

From the Implicit Function Theorem (see, e.g., Apostol (1973)), conditions (c3) and (c4), it follows that there exists a function f and $(q-r)$ subvectors θ^* and θ_0^* of θ and θ_0 respectively, such that $\theta = f(\theta^*)$, and $\theta_0 = f(\theta_0^*)$. Obviously, the optimization problem of minimizing $F(\theta)$ subject to $h(\theta) = 0$ is equivalent to minimizing $F(f(\theta^*))$ with respect to θ^*.

Let $\hat{\theta}^*$ be the minimum of $F(f(\theta^*))$, it follows that $\tilde{\theta} = f(\hat{\theta}^*)$. Let 2Φ be the asymptotic covariance of $\hat{\theta}^*$. From the Delta Theorem and Theorem 1, we have

$$H_G^{\frac{1}{2}}(\theta_0)(\tilde{\theta} - \theta_0) \xrightarrow{L} N[0, \lim_{G\to\infty} 2H_G^{\frac{1}{2}}(\theta_0)\dot{f}(\theta_0^*)'\Phi\dot{f}(\theta_0^*)H_G^{\frac{1}{2}}(\theta_0)]. \tag{2.3}$$

Sometimes it is difficult or impossible to obtain $f(.)$. The following theorem gives a more convenient expression for the asymptotic covariance matrix of $\tilde{\theta}$ that does not depend on $f(.)$.

Let $Q(\theta, \beta) = F(\theta) + \beta' h(\theta)$, where β is the vector of Lagrange multipliers, and let $\tilde{\beta}$ be its estimate. It follows from the first order necessary condition that $\dot{Q}(\tilde{\theta}, \tilde{\beta}) = 0, \quad h(\tilde{\theta}) = 0$, where $\dot{Q}(\theta, \beta) = \partial Q(\theta, \beta)/\partial\theta$.

Theorem 2. *The joint asymptotic distribution of* $H_G^{\frac{1}{2}}(\theta_0)(\tilde{\theta} - \theta_0)$ *and* $G^{-\frac{1}{2}}\tilde{\beta}$ *is* $N(0, 2B)$, *where*

$$B = \lim_{G\to\infty} \begin{pmatrix} I_q - H_G^{-\frac{1}{2}}\dot{h}(\dot{h}'H_G^{-1}\dot{h})^{-1}\dot{h}'H_G^{-\frac{1}{2}} & 0 \\ 0 & G^{-1}(\dot{h}'H_G^{-1}\dot{h})^{-1} \end{pmatrix},$$

with $h = h(\theta_0)$ *and* $\dot{h} = (\partial h(\theta)/\partial\theta)|_{\theta=\theta_0}$.

It is clear from Theorem 2 that $H_G^{\frac{1}{2}}(\tilde{\theta} - \theta_0)$ and $G^{-\frac{1}{2}}\tilde{\beta}$ are asymptotically independent. The following theorem gives us the asymptotic distribution of the goodness of fit test statistic about the constraints.

Theorem 3. *The asymptotic distribution of* $F(\tilde{\theta}) - F(\hat{\theta})$ *is chi-square with degrees of freedom* r.

Suppose $h^*(\theta) = (h_1(\theta), \ldots, h_j(\theta))$ is a subvector of $h(\theta)$ with $j < r$. Let $\tilde{\theta}^*$ be the constrainted ML estimate of θ_0 subject to $h^*(\theta) = 0$ only. Using the technique in the proof of Theorem 3, we can obtain the following corollary.

Corollary 3. *The asymptotic distribution of* $F(\tilde{\theta}) - F(\tilde{\theta}^*)$ *is chi-square with r-j degrees of freedom.*

3. An example

To provide an example with a large number of groups and small sample sizes in some of the groups, we analyze a portion of the data in a study of the relationship between AIDS and the use of condom (Morisky, Tiglao, Sneed, Tempongko, Baltazas, Detels and Stein, (1998)). The data were collected from female sexworkers in 97 establishments (nightclubs) in Philippine cities. The questionnaire involves knowledge about AIDS; belief, attitudes and behaviors; attitudes towards condoms; alcohol and drug use; etc. This is a two-level data set with establishments at the group level and sexworkers at the individual level. To illustrate the constrained ML estimation results, six variables were selected from a total of 137. The questions corresponding to the first three variables are: how great are the risks of getting AIDS or the AIDS virus from (1) kissing a person with the AIDS virus on the cheek? (2) deep kissing with someone who has the AIDS virus? and (3) having sexual intercourse with someone who has the AIDS virus using a condom? The questions corresponding to the last three variables are: (4) how much of a threat do you think AIDS is to the health of people? (5) what are the chances that you yourself might get AIDS? (6) how worried are you about getting AIDS? For brevity, observations with missing entries were deleted. The different sample sizes at the individual level are given in Table 1. The "Frequency" rows give the numbers of establishments with the corresponding sample sizes. For example, there are 6 establishments with sample sizes equal to 1, 3 establishments with sample sizes equal to 11, and so on. The total sample size is 758. There are 97 groups and some of the group sizes are small.

Table 1. The distribution of N_g in the AIDS data.

N_g	1	2	3	4	5	6	7	8	9	10	Subtotal
Frequency	6	11	13	6	5	6	11	7	7	2	74
N_g	11	12	13	15	16	17	19	28	59		Subtotal
Frequency	3	6	3	2	3	2	1	2	1		23

The raw data were standardized. Based on some exploratory analysis, the data set was analyzed via the following two-level confirmatory factor analysis model with invariant within-group covariance structure:

$$\Omega_B = \Lambda_B \Phi_B \Lambda_B' + \Psi_B; \quad \Omega_{gw} = \Omega_w = \Lambda_w \Phi_w \Lambda_w' + \Psi_w, g = 1, \ldots, G, \qquad (3.1)$$

where Λ_B and Λ_w are the factor loading matrices, Φ_B and Φ_w are the covariance matrices of the factors, and Ψ_B and Ψ_w are the diagonal convariance matrices of the error measurements, respectively. Moreover, the following specifications are imposed on the parameter matrices in the estimation:

$$\Lambda_B = [\Lambda_B(1,1) \quad \Lambda_B(2,1) \quad \Lambda_B(3,1) \quad 0^* \quad \Lambda_B(5,1) \quad 0^*]', \quad \Phi_B = 1.0^*,$$
$$\Psi_B = \mathrm{diag}(\Psi_B(1,1), \ldots, \Psi_B(6,6))$$

$$\Lambda_w = \begin{bmatrix} \Lambda_w(1,1) & \Lambda_w(2,1) & \Lambda_w(3,1) & 0^* & 0^* & 0* \\ 0^* & 0^* & 0* & \Lambda_w(4,2) & \Lambda_w(5,2) & \Lambda_w(6,2) \end{bmatrix}',$$

$$\Phi_w = \begin{bmatrix} 1.0^* & \Phi_w(2,1) \\ \Phi_w(2,1) & 1.0^* \end{bmatrix}, \quad \Psi_w = \mathrm{diag}(\Psi_w(1,1), \ldots, \Psi_w(6,6)),$$

where parameters with an "$*$" are treated as fixed parameters. Thus, the total number of unknown parameters is 23. For the sake of illustration, the following linear and nonlinear constraints are imposed:

$$\text{(i)} \quad \Lambda_B^2(k,1) + \Psi_B(k,k) + \Lambda_w^2(k,1) + \Psi_w(k,k) = 1.0, \quad \text{for } k = 1,2,3;$$

$$\text{(ii)} \quad \Lambda_B^2(k,1) + \Psi_B(k,k) + \Lambda_w^2(k,2) + \Psi_w(k,k) = 1.0, \quad \text{for } k = 4,5,6;$$

$$\text{(iii)} \quad \Lambda_w(3,1) = \frac{1}{3}\Lambda_w(1,1) = \frac{1}{4}\Lambda_w(2,1); \quad \Lambda_w(4,2) = \Lambda_w(5,2) = \Lambda_w(6,2). \tag{3.2}$$

The six nonlinear constraints were used to fix the diagonal of the covariance matrix of the manifest variables to 1.0, while the four linear constraints specify some relationships among the parameters in the loading matrix of the within-group structure. According to the basic definition of the factor analysis model (Jöreskog (1969); Mulaik (1972)), the relative sizes of the factor loadings as specified in (iii) represent corresponding effects of the manifest variables on the latent factors. These constraints are realistic and common in confirmatory factor analysis. The unconstrained and the constrained ML estimates of the unknown parameters were obtained via a modified EM algorithm as developed in Lee and Tsang (1999). The results are presented in Table 2. Note that the constrained ML estimates satisfy the constraints given in (3.2). Moreover, it can be seen from this table that there are two non-overlapping factors in the invariant within-group covariance structures. The first can be interpreted as a factor about the risk of getting AIDS, the other as a factor about the worry of AIDS. The between-group covariance structure is a single factor analysis model which basically describes the influence of the establishments. The discrepancy function $F(\theta)$ evaluated at the saturated model, Model (3.1) without any constraints, and Model (3.1) with constraints (3.2) is equal to 3857.8, 3891.5 and 3894.2, respectively. The value of the goodness-of-fit statistic for Model (3.1) is about 33.7. On 19 degrees of freedom, it seems that Model (3.1) barely fits the sample data. The goodness of fit statistic, $F(\tilde{\theta}) - F(\hat{\theta})$, for testing the constraints is about 2.7. On 10 degrees of freedom, it can be concluded that these constraints are not rejected.

Some simulation studies have been conducted to study the empirical performance of the constrained ML estimates. The results obtained indicate that the ML estimates and their standard errors estimates are accurate, and that the empirical distributions agree with the asymptotic theory developed.

Table 2. ML estimates of the AIDS data.

Parameter	Unconstrained Estimate	Constrained Estimate	Parameter	Unconstrained Estimate	Constrained Estimate
$\Lambda_B(1,1)$	0.279	0.265	$\Lambda_w(1,1)$	0.579	0.590
$\Lambda_B(2,1)$	0.328	0.316	$\Lambda_w(2,1)$	0.811	0.784
$\Lambda_B(3,1)$	0.200	0.196	$\Lambda_w(3,1)$	0.203	0.196
$\Lambda_B(5,1)$	-0.323	-0.329	$\Lambda_w(4,2)$	0.461	0.370
$\Psi_B(1,1)$	0.017	0.018	$\Lambda_w(5,2)$	0.300	0.370
$\Psi_B(2,2)$	0.026	0.025	$\Lambda_w(6,2)$	0.361	0.370
$\Psi_B(3,3)$	0.045	0.044	$\Psi_w(1,1)$	0.581	0.560
$\Psi_B(4,4)$	0.089	0.090	$\Psi_w(2,2)$	0.227	0.258
$\Psi_B(5,5)$	0.142	0.135	$\Psi_w(3,3)$	0.880	0.876
$\Psi_B(6,6)$	0.167	0.165	$\Psi_w(4,4)$	0.704	0.770
			$\Psi_w(5,5)$	0.657	0.619
			$\Psi_w(6,6)$	0.708	0.697
			$\Phi_w(2,1)$	0.259	0.250

4. Discussion

Consider the two-level structural equation models with large sample size N_g in each group; as pointed out by a reviewer (see also Lee and Poon (1993)), the statistical analysis of the within-group structures can be based on the marginal sample covariance matrices $S_g,\ g = 1, \ldots, G$. Then, since $S_1, \ldots, S_G$ are independent, analysis of this kind reduces to the standard multisample analysis of single-level models with constraints. Statistical theory for the latter is well established; see, Lee and Tsui (1982), EQS (Bentler (1992)) and LISREL (Jöreskog and Sörbom (1996)). However, many practical problems involve small N_g. These situations are less straightforward because we cannot use the asymptotic properties of S_g and $\bar{x}_g$ directly to establish the results. To analyze the between-group structure and to achieve the asymptotic properties, we require a significantly large number of groups to provide the required information.

Let $F^+(\theta)$ be the GLS function given in Lee (1990), $\dot{F}^+(\theta)$ and $\ddot{F}^+(\theta)$ be the corresponding gradient vector and the Hessian matrix, respectively, and $\hat{\theta}^+$ and $\tilde{\theta}^+$ be the unconstrained and the constrained GLS estimators of θ_0, respectively. It can be shown that lemmas, corollaries, and theorems about ML estimation presented in Section 2 are also valid for the GLS estimation if $F(\theta)$, $\dot{F}(\theta)$, $\ddot{F}(\theta)$, $\hat{\theta}$ and $\tilde{\theta}$ are replaced by $F^+(\theta)$, $\dot{F}^+(\theta)$, $\ddot{F}^+(\theta)$, $\hat{\theta}^+$ and $\tilde{\theta}^+$, respectively.

Acknowledgement

This research was supported by a HKUGC Earmarked grant 4026/97H. The authors are deeply indebted to an associate editor and two anonymous reviewers for variable comments, and to D. E. Moriskey and J. A. Stein for the use of their AIDS data. The assistance of Michael K. H. Leung and Esther L. S. Tam is also gratefully acknowledged.

Appendix. Proofs of theorems

Proof of Theorem 1. From the Mean Value Theorem, there exists a θ_* between θ_0 and $\hat{\theta}$ such that $-H_G^{-\frac{1}{2}}\dot{F}(\theta_0) = H_G^{-\frac{1}{2}}\ddot{F}(\theta_*)H_G^{-\frac{1}{2}}H_G^{\frac{1}{2}}(\hat{\theta}-\theta_0)$. From Corollary 2 and Lemma 2, we have $H_G^{-\frac{1}{2}}\ddot{F}(\theta_*)H_G^{-\frac{1}{2}} \xrightarrow{P} I_q$. So

$$H_G^{\frac{1}{2}}(\hat{\theta}-\theta_0) + H_G^{-\frac{1}{2}}\dot{F}(\theta_0) \xrightarrow{P} 0. \tag{A.1}$$

It follows from Lemma 3 that $H_G^{\frac{1}{2}}(\hat{\theta}-\theta_0) \xrightarrow{L} N(0, 2I_q)$.

Proof of Theorem 2. From the Mean Value Theorem, there exits a θ_* between θ_0 and $\tilde{\theta}$, such that

$$-\dot{F}(\theta_0) = \ddot{Q}(\theta_*, \tilde{\beta})(\tilde{\theta}-\theta_0) + \dot{h}\tilde{\beta} = \ddot{F}(\theta_*)(\tilde{\theta}-\theta_0) + \dot{h}\tilde{\beta} + \ddot{h}(\theta_*)(I_q \otimes \tilde{\beta})(\tilde{\theta}-\theta_0); \text{and}$$

$$-H_G^{-\frac{1}{2}}\dot{F}(\theta_0) = H_G^{-\frac{1}{2}}\ddot{F}(\theta_*)H_G^{-\frac{1}{2}}H_G^{\frac{1}{2}}(\tilde{\theta}-\theta_0) + H_G^{-\frac{1}{2}}\dot{h}\tilde{\beta} + H_G^{-\frac{1}{2}}\ddot{h}(\theta_*)(I_q \otimes \tilde{\beta})(\tilde{\theta}-\theta_0). \tag{A.2}$$

From Corollary 2, (2.3) and Theorem 1, we have

$$\|H_G^{-\frac{1}{2}}\ddot{F}(\theta_*)H_G^{-\frac{1}{2}} - I_q\| \xrightarrow{P} 0. \tag{A.3}$$

Because $\dot{F}(\tilde{\theta}) + \dot{h}(\tilde{\theta})\tilde{\beta} = 0$, we have $G^{-1}\tilde{\beta} = -\left[\dot{h}(\tilde{\theta})'\dot{h}(\tilde{\theta})\right]^{-1}\dot{h}(\tilde{\theta})'\left[G^{-1}\dot{F}(\tilde{\theta})\right]$. Similarly, from Corollary 2, Lemma 3 and Theorem 1, we have $G^{-1}[\dot{F}(\tilde{\theta}) - \dot{F}(\theta_0)] \xrightarrow{P} 0$, and $G^{-1}\dot{F}(\tilde{\theta}) \xrightarrow{P} 0$. So $G^{-1}\tilde{\beta}$ converges to zero in probability. Because $h(\tilde{\theta}) = h(\theta_0) = 0$, there exits a θ_+ between θ_0 and $\tilde{\theta}$, such that $0 =$

$G^{\frac{1}{2}}[h(\tilde{\theta}) - h] = G^{\frac{1}{2}}\dot{h}(\theta_+)'H_G^{-\frac{1}{2}}H_G^{\frac{1}{2}}(\tilde{\theta} - \theta_0)$. From (A.2), (A.3) and Theorem 1, it can be shown that

$$\begin{pmatrix} H_G^{\frac{1}{2}}(\tilde{\theta} - \theta_0) \\ G^{-\frac{1}{2}}\tilde{\beta} \end{pmatrix} + \begin{pmatrix} I_q - H_G^{-\frac{1}{2}}\dot{h}(\dot{h}'H_G^{-1}\dot{h})^{-1}\dot{h}'H_G^{-\frac{1}{2}} \\ G^{-\frac{1}{2}}(\dot{h}'H_G^{-1}\dot{h})^{-1}\dot{h}'H_G^{-\frac{1}{2}} \end{pmatrix} H_G^{-\frac{1}{2}}\dot{F}(\theta_0) \xrightarrow{P} 0. \quad \text{(A.4)}$$

On the basis of the result in Lemma 3, the proof is complete.

Proof of Theorem 3. There exists a θ_* between θ_0 and $\tilde{\theta}$, and a θ_{**} between θ_0 and $\hat{\theta}$ such that $F(\theta_0) - F(\tilde{\theta}) = \frac{1}{2}(\theta_0 - \tilde{\theta})'\ddot{Q}(\theta_*, \tilde{\beta})(\theta_0 - \tilde{\theta})$, and $F(\theta_0) - F(\hat{\theta}) = \dot{F}(\hat{\theta}) + \frac{1}{2}(\theta_0 - \hat{\theta})'\ddot{F}(\theta_{**})(\theta_0 - \hat{\theta})$. Since $\dot{F}(\hat{\theta}) = 0$, we can obtain via Corollary 2, Lemma 2 and Theorem 1 that

$$F(\tilde{\theta}) - F(\hat{\theta}) - \{\frac{1}{2}(H_G^{\frac{1}{2}}(\theta_0 - \hat{\theta}))'(H_G^{\frac{1}{2}}(\theta_0 - \hat{\theta})) - \frac{1}{2}(H_G^{\frac{1}{2}}(\theta_0 - \tilde{\theta}))'(H_G^{\frac{1}{2}}(\theta_0 - \tilde{\theta}))\} \xrightarrow{P} 0. \quad \text{(A.5)}$$

From (A.4), we have $H_G^{\frac{1}{2}}(\tilde{\theta} - \theta_0) + (I_q - H_G^{-\frac{1}{2}}\dot{h}(\dot{h}'H_G^{-1}\dot{h})^{-1}\dot{h}'H_G^{-\frac{1}{2}})H_G^{-\frac{1}{2}}\dot{F}(\theta_0) \xrightarrow{P} 0$. Using (A.1),

$$\begin{aligned} & F(\tilde{\theta}) - F(\hat{\theta}) - 2^{-1}\{(H_G^{-\frac{1}{2}}\dot{F}(\theta_0))'H_G^{-\frac{1}{2}}\dot{F}(\theta_0) \\ & -(H_G^{-\frac{1}{2}}\dot{F}(\theta_0))'[I_q - H_G^{-\frac{1}{2}}\dot{h}(\dot{h}'H_G^{-1}\dot{h})^{-1}\dot{h}'H_G^{-\frac{1}{2}}]^2 H_G^{-\frac{1}{2}}\dot{F}(\theta_0)\} \\ = & F(\tilde{\theta}) - F(\hat{\theta}) - [2^{-\frac{1}{2}}H_G^{-\frac{1}{2}}\dot{F}(\theta_0)]'[H_G^{-\frac{1}{2}}\dot{h}(\dot{h}'H_G^{-1}\dot{h})^{-1}\dot{h}'H_G^{-\frac{1}{2}}] \\ & \cdot[2^{-\frac{1}{2}}H_G^{-\frac{1}{2}}\dot{F}(\theta_0)] \xrightarrow{P} 0. \end{aligned} \quad \text{(A.6)}$$

Since $H_G^{-\frac{1}{2}}\dot{h}(\dot{h}'H_G^{-1}\dot{h})^{-1}\dot{h}'H_G^{-\frac{1}{2}}$ is an idempotent matrix with rank r, it follows from (A.6) that $F(\tilde{\theta}) - F(\hat{\theta}) \xrightarrow{L} \chi^2_r$.

References

Apostol, T. M. (1973). *Mathematical Analysis.* 2nd edition. Addison-Wesley, Reading, Mass.

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

Bentler, P. M. (1992). *EQS Structural Equations Program Manual.* BMDP Statistical Software, Los Angeles.

Cuttance, P. and Ecob, R. (1987). *Structural Modeling by Example.* Cambridge University Press, Cambridge.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34**, 183-202.

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

Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language.* Scientific Software International, London.

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

Lee, S. Y. (1990). Multilevel analysis of structural equation models. *Biometrika* **77** 763-772.

Lee, S. Y. and Poon, W. Y. (1992). Two level analysis of covariance structures for unbalanced designs with small level-one samples. *British J. Math. Statist. Psych.* **45**, 109-123.

Lee, S. Y. and Poon, W. Y. (1993). Structural equation models with hierarchical data. In *Statist. Model. Latent Variable* (Edited by K. Haagan, D. J. Bartholomew and M. Deistler), 203-228. North-Holland.

Lee, S. Y. and Poon, W. Y. (1998). Analysis of two-level structural equation models via EM type algorithms. *Statist. Sinica* **8**, 749-766.

Lee, S. Y. and Tsang, S. Y. (1999). Constrained maximum likelihood estimation of two-level covariance structure model via EM type algorithms. *Psychometrika* **64**, 435-450.

Lee, S. Y. and Tsui, K. L. (1982). Covariance structure analysis in several populations. *Psychometrika* **47**, 297-308.

Longford, N. T. and Muthén, B. O. (1992). Factor analysis for clustered observations. *Psychometrika* **57**, 581-597.

McDonald, R. P. and Goldstein, H. I. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British J. Math. Statist. Psych.* **42**, 215-232.

Moriskey, D. E., Tiglao, T. V., Sneed, C. D., Tempongko, S. B., Baltazas, J. C., Detels, R. and Stein, J. A (1998). The effects of establishments practices, knowledge and attitudes on condom use among Filipina sex workers. *AIDS Case* **10**, 213-220.

Mulaik, S. A. (1972). Factor analysis and Psychometrika. *Psychometrika* **51**, 23-34.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika* **54**, 557-585.

Department of Statistics, The London School of Economics and Political Science, U.K.

Department of Statistics, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong.
E-mail: sylee@sparc2.sta.cuhk.edu.hk

(Received January 1998; accepted May 2000)