Statistica Sinica **14**(2004), 317-332

LOCAL INFLUENCE ANALYSIS OF TWO-LEVEL LATENT VARIABLE MODELS WITH CONTINUOUS AND POLYTOMOUS DATA

Xin-Yuan Song[1,2] and Sik-Yum Lee[1]

[1] *The Chinese University of Hong Kong and* [2] *Zhongshan University*

Abstract: A latent variable model is proposed to analyze two-level data with hierarchical structure and mixed continuous and polytomous data that are very common in behavioral, biomedical and social research. On the basis of an EM algorithm associated with the maximum likelihood estimation of the model, a method is developed for assessing local influence of minor perturbation for the proposed latent variable model. The key idea of the development is to derive diagnostic measures on the basis of the conditional expectation of the complete-data log-likelihood function in the E-step of the EM algorithm. Building blocks in the diagnostic measures are computed via observations generated by the Gibbs sampler. It is shown that the proposed method is computationally efficient and feasible for a wide variety of perturbations that carry clear interpretation. The approach is illustrated by a two-level data set concerning the development and findings from an AIDS preventive intervention of Filipino commercial sex workers.

Key words and phrases: Benchmark, conditional expectation, Gibbs sampler, MCEM algorithm, perturbation.

1. Introduction

In behavioral, medical and social research, substantive theory involves latent variables that cannot be assessed by a single measurement on each individual under study. Relationships among latent and manifest variables are important in establishing a model for making correct decisions. Latent variable models such as the LISREL model (Jöreskog and Sörbom (1996)) and Sammel and Ryan's (1996) model are important in analyzing these relations, and have been extensively applied. While the above models are developed for continuous variables, data are often measured on ordinal response scales. These measurements are usually modeled by polytomous variables that are defined through underlying continuous variables with thresholds. Recently, a lot of attention has been focused on latent variable models with mixed polytomous and continuous variables, see Sammel, Ryan and Legler (1997), Shi and Lee (2000) and Song and Lee (2001), among others. Methods developed in this work depend on the assumption of independence among observations. However, it is common to encounter hierarchically structured data that are collected from units that are nested within a large cluster. For multilevel data, the assumption of independence is not realistic because individuals within a cluster are expected to share certain common influential factors. In analysis of latent variable models, the need to develop statistical methods for multilevel data is well recognized; see for example, Lee and Shi (2001). However, existing development of two-level latent variable models has only focused on estimation.

Local influence analysis is a general statistical technique to assess the stability of estimation outputs with respect to models inputs. Cook (1986) proposed a unified approach for assessment of local influence in minor perturbations of a statistical model. In past years, this powerful approach dominated the local influence analysis of statistical models (see Lesaffre and Verbeke (1998), among others) and latent variable models (see Poon, Wang and Lee (1999), among others). However, no local influence analysis for two-level models with mixed continuous and polytomous variables has been developed. The main reason may be that the building blocks in the diagnostic measures associated with Cook's (1986) approach involve intractable integrals induced by the complexities of the data structures.

The main objective of this article is to develop diagnostic measures for local influence analysis of a general two-level latent variable model with mixed continuous and polytomous variables. To achieve our goal, we have to develop a Monte Carlo EM (MCEM) algorithm (Dempster, Laird and Rubin (1977); Wei and Tanner (1990)) for maximum likelihood (ML) estimation. Our development of local influence measures is on the basis of a general approach proposed by Zhu and Lee (2001). In the local influence analysis, we focus on a displacement function that depends on the conditional expectation of the complete-data log-likelihood at the E-step of the EM algorithm rather than the more complicated observed-data likelihood displacement function as in Cook's (1986) famous approach. As our procedure does not require evaluation of any intractable integrals, its theoretical development is manageable and its computation burden is not heavy.

Section 2 introduces a general two-level latent variables model (LVM) and describes briefly an MCEM type algorithm for obtaining the ML estimates. A procedure for assessing the local influence of the proposed model is developed in Section 3, where diagnostic measures are obtained via the conformal normal curvature. In Section 4, an illustrative example based on an AIDS data set is presented. A discussion is given in Section 5.

2. A General Two-level LVM with Mixed Type Variables

Consider a collection of p-variate random vectors $\mathbf{u}_{gi}$, $i = 1, \ldots, N_g$, within groups $g = 1, \ldots, G$. The sample sizes N_g may differ from group to group so that the data set is unbalanced. At the first level, we assume that, conditional on the group mean $\mathbf{v}_g$, random observations in each group have the following structure:

$$\mathbf{u}_{gi} = \mathbf{v}_g + \mathbf{v}_{gi} = \mathbf{v}_g + \mathbf{\Lambda}_w \boldsymbol{\zeta}_{gi} + \boldsymbol{\varepsilon}_{gi}, \quad g = 1, \ldots, G, \ i = 1, \ldots, N_g, \tag{1}$$

where $\mathbf{\Lambda}_w$ is a $p \times q_w$ matrix of factor loadings, $\boldsymbol{\zeta}_{gi}$ is a $q_w \times 1$ random vector of latent factors, and $\boldsymbol{\varepsilon}_{gi}$ is a $p \times 1$ random vector of error measurements which is independent of $\boldsymbol{\zeta}_{gi}$ and is distributed as $N[\mathbf{0}, \mathbf{\Psi}_w]$, where $\mathbf{\Psi}_w$ is a diagonal matrix. At the second level, we assume that the group mean $\mathbf{v}_g$ has the structure

$$\mathbf{v}_g = \boldsymbol{\mu} + \mathbf{\Lambda}_b \boldsymbol{\zeta}_g + \boldsymbol{\varepsilon}_g, \quad g = 1, \ldots, G, \tag{2}$$

where $\boldsymbol{\mu}$ is the mean vector, $\mathbf{\Lambda}_b$ is a $p \times q_b$ matrix of factor loadings, $\boldsymbol{\zeta}_g$ is a $q_b \times 1$ vector of latent variables, and $\boldsymbol{\varepsilon}_g$ is a $p \times 1$ random vector of error measurements which is independent of $\boldsymbol{\zeta}_g$ and is distributed as $N[\mathbf{0}, \mathbf{\Psi}_b]$, where $\mathbf{\Psi}_b$ is a diagonal matrix. Moreover, the first level latent vectors $\boldsymbol{\zeta}_{gi}$ and $\boldsymbol{\varepsilon}_{gi}$ are assumed to be independent of the second level latent vectors $\boldsymbol{\zeta}_g$ and $\boldsymbol{\varepsilon}_g$. To handle more complex situations, the latent vectors $\boldsymbol{\zeta}_{gi}$ and $\boldsymbol{\zeta}_g$ are partitioned as $\boldsymbol{\zeta}_{gi} = (\boldsymbol{\eta}_{gi}^T, \boldsymbol{\xi}_{gi}^T)^T$ and $\boldsymbol{\zeta}_g = (\boldsymbol{\eta}_g^T, \boldsymbol{\xi}_g^T)^T$, respectively, where $\boldsymbol{\eta}_{gi}(q_{w1} \times 1), \boldsymbol{\xi}_{gi}(q_{w2} \times 1), \boldsymbol{\eta}_g(q_{b1} \times 1)$ and $\boldsymbol{\xi}_g(q_{b2} \times 1)$ are latent vectors, with $q_{w1} + q_{w2} = q_w$, and $q_{b1} + q_{b2} = q_b$. Moreover, the two-level model involves the following structural equations:

$$\boldsymbol{\eta}_{gi} = \mathbf{\Pi}_w \boldsymbol{\eta}_{gi} + \mathbf{\Gamma}_w \boldsymbol{\xi}_{gi} + \boldsymbol{\delta}_{gi}, \quad \text{and} \quad \boldsymbol{\eta}_g = \mathbf{\Pi}_b \boldsymbol{\eta}_g + \mathbf{\Gamma}_b \boldsymbol{\xi}_g + \boldsymbol{\delta}_g, \tag{3}$$

in the between-groups and within-groups models, respectively, where $\mathbf{\Pi}_w(q_{w1} \times q_{w1}), \mathbf{\Pi}_b(q_{b1} \times q_{b1}), \mathbf{\Gamma}_w(q_{w1} \times q_{w2})$ and $\mathbf{\Gamma}_b(q_{b1} \times q_{b2})$ are unknown parameter matrices. We assume that $\mathbf{I}_w - \mathbf{\Pi}_w$ and $\mathbf{I}_b - \mathbf{\Pi}_b$ are nonsingular and their determinants are independent of $\mathbf{\Pi}_w$ and $\mathbf{\Pi}_b$. Hence, this is a general two-level model with structural equations for assessing relationships among latent variables at both the individual and group levels.

To study the model with mixed polytomous and continuous variables, let $\mathbf{u}_{gi} = (\mathbf{x}_{gi}^T, \mathbf{y}_{gi}^T)^T$, where $\mathbf{x}_{gi} = (x_{gi1}, \ldots, x_{gir})^T$ is an observable continuous random vector and $\mathbf{y}_{gi} = (y_{gi1}, \ldots, y_{gis})^T$ an unobservable continuous random vector. A probit model is used to model the observable polytomous vector $\mathbf{z} = (z_1, \ldots, z_s)^T$ with its underlying continuous vector $\mathbf{y} = (y_1, \ldots, y_s)^T$ as follows: if y_k is in $(\alpha_{k,z_k}, \alpha_{k,z_k+1}]$, then the kth entry of $\mathbf{z}$ is equal to an integral value z_k in $\{0, \ldots, b_k\}$. In general, we let $\alpha_{k,0} = -\infty$, $\alpha_{k,b_k+1} = \infty$. For the kth variable, there are $b_k + 1$ categories which are defined by unknown thresholds α_{kj}.

The proposed two-level LVM with structural equations at both levels and with mixed continuous and polytomous variables is rather general. It subsumes many existing LVMs such as the well-known LISREL model (Jöreskog and

Sörbom(1996)), and the model in Shi and Lee (2000). Although we assume, for simplicity of notation and brevity, that $\mathbf{\Lambda}_w$, $\mathbf{\Phi}_w$ and $\mathbf{\Psi}_w$ are invariant across groups, it is straightforward to extend the methodologies developed here to models without this assumption.

To identify the variance and the thresholds associated with each polytomous variable, $\alpha_{k,1}$ and α_{k,b_k}, $k = 1, \ldots, s$, are fixed at some preassigned values. This is basically equivalent to use the range $\alpha_{k,b_k} - \alpha_{k,1}$ as a measure for the dispersion of the polytomous variable z_k. The choice of the preassigned values only changes the scale of the covariance matrices but not the correlations of the latent variables, see Lee and Shi (2001). Moreover, we follow the common method in latent variable modeling to identify the within-groups and between-groups covariance structures by fixing appropriate elements in $\mathbf{\Lambda}_w$, $\mathbf{\Lambda}_b$, $\mathbf{\Pi}_w$, $\mathbf{\Gamma}_w$, $\mathbf{\Pi}_b$, $\mathbf{\Gamma}_b$, $\mathbf{\Phi}_w$, and/or $\mathbf{\Phi}_b$ at preassigned known values, see Jöreskog and Sörbom (1996). Let $\boldsymbol{\alpha}$ be a vector that contains all unknown thresholds, and $\boldsymbol{\theta}$ be the parameter vector that contains all unknown structural parameters in $\mathbf{\Lambda}_w$, $\mathbf{\Psi}_w$, $\mathbf{\Pi}_w$, $\mathbf{\Gamma}_w$, $\mathbf{\Phi}_w$, $\mathbf{\Psi}_{w\delta}$, $\mathbf{\Lambda}_b$, $\mathbf{\Psi}_b$, $\mathbf{\Pi}_b$, $\mathbf{\Gamma}_b$, $\mathbf{\Phi}_b$, $\mathbf{\Psi}_{b\delta}$ and $\boldsymbol{\alpha}$. The ML estimation is briefly outlined in the next section.

3. ML Estimation via the MCEM Algorithm

Let $\mathbf{X} = \{\mathbf{x}_{gi} : \ g = 1, \ldots, G; \ i = 1, \ldots, N_g\}$ and $\mathbf{Z} = \{\mathbf{z}_{gi} : \ g = 1, \ldots, G; \ i = 1, \ldots, N_g\}$ be the observed data matrices, and $L_o(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z})$ be the observed-data log-likelihood function. Owing to the nature of the two-level polytomous data, $L_o(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z})$ involves intractable multiple integrals. Thus, it is very difficult to obtain the ML estimate of $\boldsymbol{\theta}$ or the diagnostic measures on the basis of Cook's (1986) approach by working directly with $L_o(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z})$. To solve the difficulties, we reformulate the problem as a missing-data problem by treating some latent quantities as hypothetical missing data and obtain the ML estimates via the well-known EM algorithm. In our model, there are a tremendous number of latent quantities, namely, the continuous measurements in $\mathbf{Y} = \{\mathbf{y}_{gi} : g = 1, \ldots, G; i = 1, \ldots, N_g\}$ that underly the polytomous variables, the latent variables in $\mathbf{F}_w = \{\boldsymbol{\zeta}_{gi} : g = 1, \ldots, G; i = 1, \ldots, N_g\}$, $\mathbf{F}_b = \{\boldsymbol{\zeta}_g : g = 1, \ldots, G\}$, and $\mathbf{V} = \{\mathbf{v}_g : g = 1, \ldots, G\}$. There are two possible ways to apply the EM algorithm. The first one treats $\mathbf{F}_w$, $\mathbf{F}_b$, $\mathbf{V}$ and $\mathbf{Y}$ as hypothetical missing data, while the second one treats only $\mathbf{F}_w$, $\mathbf{F}_b$ and $\mathbf{V}$ as missing.

Let $\mathbf{\Lambda}_{w1}$ and $\mathbf{\Lambda}_{w2}$ be the sub-matrices of $\mathbf{\Lambda}_w$, $\mathbf{\Psi}_{w1}$ and $\mathbf{\Psi}_{w2}$ be the sub-matrices of $\mathbf{\Psi}_w$, and $\mathbf{v}_{g1}$ and $\mathbf{v}_{g2}$ be the sub-vectors of $\mathbf{v}_g$ that are associated with $\mathbf{x}$ and $\mathbf{y}$, respectively. For $h = 1, 2$, let $\mathbf{\Lambda}_{whk}$, ψ_{whk} and v_{ghk} be the kth row, kth diagonal element and kth element of $\mathbf{\Lambda}_{wh}$, $\mathbf{\Psi}_{wh}$ and $\mathbf{v}_{gh}$, respectively. In the first EM algorithm, the complete-data log-likelihood function,

$L_c^*(\boldsymbol{\theta}|\mathbf{X},\mathbf{Z},\mathbf{Y},\mathbf{F}_w,\mathbf{F}_b,\mathbf{V})$, is proportional to the sums of the following functions:

$$L_1^*(\boldsymbol{\Lambda}_{w1},\boldsymbol{\Psi}_{w1}|\mathbf{X},\mathbf{F}_w,\mathbf{V}) = \sum_{g=1}^{G}\sum_{i=1}^{N_g}\log p(\mathbf{x}_{gi}|\boldsymbol{\zeta}_{gi},\mathbf{v}_{g1},\boldsymbol{\Lambda}_{w1},\boldsymbol{\Psi}_{w1})$$

$$= \sum_{g=1}^{G}\sum_{i=1}^{N_g} -\frac{1}{2}\left\{\log|\boldsymbol{\Psi}_{w1}| + (\mathbf{x}_{gi}-\mathbf{v}_{g1}-\boldsymbol{\Lambda}_{w1}\boldsymbol{\zeta}_{gi})^T\boldsymbol{\Psi}_{w1}^{-1}(\mathbf{x}_{gi}-\mathbf{v}_{g1}-\boldsymbol{\Lambda}_{w1}\boldsymbol{\zeta}_{gi})\right\}, \quad (4)$$

$$L_2^*(\boldsymbol{\Lambda}_{w2},\boldsymbol{\Psi}_{w2},\boldsymbol{\alpha}|\mathbf{Z},\mathbf{Y},\mathbf{F}_w,\mathbf{V}) = \sum_{g=1}^{G}\sum_{i=1}^{N_g}\sum_{k=1}^{s}\log I_{(\alpha_{k,z_{gik}},\alpha_{k,z_{gik}+1}]}(y_{gik})$$

$$-\frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{N_g}\left\{\log|\boldsymbol{\Psi}_{w2}| + (\mathbf{y}_{gi}-\mathbf{v}_{g2}-\boldsymbol{\Lambda}_{w2}\boldsymbol{\zeta}_{gi})^T\boldsymbol{\Psi}_{w2}^{-1}(\mathbf{y}_{gi}-\mathbf{v}_{g2}-\boldsymbol{\Lambda}_{w2}\boldsymbol{\zeta}_{gi})\right\},$$

$$L_3^*(\boldsymbol{\Pi}_w,\boldsymbol{\Gamma}_w,\boldsymbol{\Phi}_w,\boldsymbol{\Psi}_{w\delta}|\mathbf{F}_w) = \sum_{g=1}^{G}\sum_{i=1}^{N_g}\log p(\boldsymbol{\zeta}_{gi}|\boldsymbol{\Pi}_w,\boldsymbol{\Gamma}_w,\boldsymbol{\Phi}_w,\boldsymbol{\Psi}_{w\delta})$$

$$= \sum_{g=1}^{G}\sum_{i=1}^{N_g} -\frac{1}{2}\Big\{\log|\boldsymbol{\Psi}_{w\delta}| + \log|\boldsymbol{\Phi}_w| + \boldsymbol{\xi}_{gi}^T\boldsymbol{\Phi}_w^{-1}\boldsymbol{\xi}_{gi}$$
$$+(\boldsymbol{\eta}_{gi}-\boldsymbol{\Pi}_w\boldsymbol{\eta}_{gi}-\boldsymbol{\Gamma}_w\boldsymbol{\xi}_{gi})^T\boldsymbol{\Psi}_{w\delta}^{-1}(\boldsymbol{\eta}_{gi}-\boldsymbol{\Pi}_w\boldsymbol{\eta}_{gi}-\boldsymbol{\Gamma}_w\boldsymbol{\xi}_{gi})\Big\}, \quad (5)$$

$$L_4^*(\boldsymbol{\mu},\boldsymbol{\Lambda}_b,\boldsymbol{\Psi}_b|\mathbf{F}_b,\mathbf{V}) = \sum_{g=1}^{G}\log p(\mathbf{v}_g|\boldsymbol{\zeta}_g,\boldsymbol{\mu},\boldsymbol{\Lambda}_b,\boldsymbol{\Psi}_b)$$

$$= \sum_{g=1}^{G} -\frac{1}{2}\left\{\log|\boldsymbol{\Psi}_b| + (\mathbf{v}_g-\boldsymbol{\mu}-\boldsymbol{\Lambda}_b\boldsymbol{\zeta}_g)^T\boldsymbol{\Psi}_b^{-1}(\mathbf{v}_g-\boldsymbol{\mu}-\boldsymbol{\Lambda}_b\boldsymbol{\zeta}_g)\right\}, \quad (6)$$

$$L_5^*(\boldsymbol{\Pi}_b,\boldsymbol{\Gamma}_b,\boldsymbol{\Phi}_b,\boldsymbol{\Psi}_{b\delta}|\mathbf{F}_b) = \sum_{g=1}^{G}\log p(\boldsymbol{\zeta}_g|\boldsymbol{\Pi}_b,\boldsymbol{\Gamma}_b,\boldsymbol{\Phi}_b,\boldsymbol{\Psi}_{b\delta})$$

$$= \sum_{g=1}^{G} -\frac{1}{2}\Big\{\log|\boldsymbol{\Psi}_{b\delta}| + \log|\boldsymbol{\Phi}_b| + \boldsymbol{\xi}_g^T\boldsymbol{\Phi}_b^{-1}\boldsymbol{\xi}_g$$
$$+(\boldsymbol{\eta}_g-\boldsymbol{\Pi}_b\boldsymbol{\eta}_g-\boldsymbol{\Gamma}_b\boldsymbol{\xi}_g)^T\boldsymbol{\Psi}_{b\delta}^{-1}(\boldsymbol{\eta}_g-\boldsymbol{\Pi}_b\boldsymbol{\eta}_g-\boldsymbol{\Gamma}_b\boldsymbol{\xi}_g)\Big\}, \quad (7)$$

where $I_A(y)$ is an indicator function which takes the value 1 if $y \in A$ and zero otherwise. The ML estimate of $\boldsymbol{\theta}$ can be obtained via the first EM algorithm which is implemented as follows at the jth iteration with current value $\boldsymbol{\theta}^{(j)}$. E-step: evaluate $Q^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) = E\{L_c^*(\boldsymbol{\theta}|\mathbf{X},\mathbf{Z},\mathbf{Y},\mathbf{F}_w,\mathbf{F}_b,\mathbf{V})|\mathbf{X},\mathbf{Z},\boldsymbol{\theta}^{(j)}\}$, where the expectation is taken with respect to the conditional distribution of $(\mathbf{Y},\mathbf{F}_w,\mathbf{F}_b,\mathbf{V})$ given $(\mathbf{X},\mathbf{Z})$ and $\boldsymbol{\theta}^{(j)}$; M-step: obtain a new value of $\boldsymbol{\theta}$ by maximizing $Q^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)})$. We apply the Gibbs sampler (Geman and Geman (1984)) to generate a sufficiently large number, say J, of observations from $p(\mathbf{Y},\mathbf{F}_w,\mathbf{F}_b,\mathbf{V}|\mathbf{X},\mathbf{Z},\boldsymbol{\theta}^{(j)})$ for computing the conditional expectation at the E-step. The choice of J proceeds on a problem-by-problem basis; for example, we can take $J = 50 + 10r$ at the rth iterations of the MCEM algorithm. The M-step is completed by approximating the various sufficient statistics via the means of the random observations generated in the E-step. We monitor convergence of the algorithm by bridge sampling (Meng and Wong (1996)), via a slight extension of the procedure described in Lee and Shi (2001, p.790).

The first EM algorithm is rather efficient for producing ML estimates. However, as $L_c^*(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z}, \mathbf{Y}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})$ is not differentiable with respect to $\boldsymbol{\alpha}$, we develop our local influence analysis on the basis of the second EM algorithm. Now the complete-data set is $(\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})$, and the complete-data log-likelihood function is equal to

$$\begin{aligned} L_c(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V}) &= L_1(\boldsymbol{\Lambda}_{w1}, \boldsymbol{\Psi}_{w1}|\mathbf{X}, \mathbf{F}_w, \mathbf{V}) + L_2(\boldsymbol{\Lambda}_{w2}, \boldsymbol{\Psi}_{w2}, \boldsymbol{\alpha}|\mathbf{Z}, \mathbf{F}_w, \mathbf{V}) \\ &+ L_3(\boldsymbol{\Pi}_w, \boldsymbol{\Gamma}_w, \boldsymbol{\Phi}_w, \boldsymbol{\Psi}_{w\delta}|\mathbf{F}_w) + L_4(\boldsymbol{\mu}, \boldsymbol{\Lambda}_b, \boldsymbol{\Psi}_b|\mathbf{F}_b, \mathbf{V}) + L_5(\boldsymbol{\Pi}_b, \boldsymbol{\Gamma}_b, \boldsymbol{\Phi}_b, \boldsymbol{\Psi}_{b\delta}|\mathbf{F}_b), \end{aligned} \quad (8)$$

where the $L_h = L_h^*$ are defined by equations (4)-(7), for $h = 1, 3, 4, 5$, while

$$L_2(\boldsymbol{\Lambda}_{w2}, \boldsymbol{\Psi}_{w2}, \boldsymbol{\alpha}|\mathbf{Z}, \mathbf{F}_w, \mathbf{V}) = \sum_{g=1}^{G} \sum_{i=1}^{N_g} \log p(\mathbf{z}_{gi}|\boldsymbol{\zeta}_{gi}, \mathbf{v}_{g2}, \boldsymbol{\Lambda}_{w2}, \boldsymbol{\Psi}_{w2}, \boldsymbol{\alpha}),$$

with $p(\mathbf{z}_{gi}|\boldsymbol{\zeta}_{gi}, \mathbf{v}_{g2}, \boldsymbol{\Lambda}_{w2}, \boldsymbol{\Psi}_{w2}, \boldsymbol{\alpha}) = \prod_{k=1}^{s}[\Phi(\alpha^*_{k,z_{gik}+1}) - \Phi(\alpha^*_{k,z_{gik}})]$, in which $\alpha^*_{k,z_{gik}} = \psi_{w2k}^{-1/2}\,(\alpha_{k,z_{gik}} - v_{g2k} - \boldsymbol{\Lambda}_{w2k}\boldsymbol{\zeta}_{gi})$, and Φ is the cumulative distribution function of $N[0,1]$. Note that this complete-data log-likelihood function contains separate terms that involve different separable parameters. Hence, its Hessian matrix is a diagonal block matrix. This reduces the computational effort for computing the influence diagnostics.

The E-step of the EM algorithm evaluates $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) = E\{L_c(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})\ |\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}^{(j)}\}$. It is completed as before. At the M-step, optimal values of $\boldsymbol{\Lambda}_{w1}$, $\boldsymbol{\Psi}_{w1}$, $\boldsymbol{\Pi}_w$, $\boldsymbol{\Phi}_w$, $\boldsymbol{\Psi}_{w\delta}$, $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}_b$, $\boldsymbol{\Psi}_b$, $\boldsymbol{\Pi}_b$, $\boldsymbol{\Phi}_b$ and $\boldsymbol{\Psi}_{b\delta}$ can be obtained in closed form. However, one requires an iterative procedure such as the Newton Raphson algorithm to obtain the optimal values of $\boldsymbol{\Lambda}_{w2}$, $\boldsymbol{\Psi}_{w2}$ and $\boldsymbol{\alpha}$ in L_2.

4. Local Inference of the Model

Let $L_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})$ be the perturbed complete-data log-likelihood function with respect to a perturbation vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^T$. Assume that there is a $\boldsymbol{\omega}^0$ such that $L_c(\boldsymbol{\theta}, \boldsymbol{\omega}^0|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V}) = L_c(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})$ for all $\boldsymbol{\theta}$. Our local influence approach is based on the Q-displacement function $f_Q(\boldsymbol{\omega}) = 2\{Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\hat{\boldsymbol{\theta}})\}$, where $\hat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ is the estimate of $\boldsymbol{\theta}$ which maximizes $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = E\{L_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})|\mathbf{X}, \mathbf{Z}, \hat{\boldsymbol{\theta}}\}$. The main motivation for using the Q-displacement function $f_Q(\boldsymbol{\omega})$ instead of the likelihood displacement function is that $L_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})$ is much simpler than $L_o(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z})$. Moreover, it can be regarded as a measure of the difference between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})$: it is greater than or equal to zero and achieves its global minimum at $\boldsymbol{\omega}^0$. When no perturbation is introduced, $\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}^0)$ is $\hat{\boldsymbol{\theta}}$. Following Cook (1986), attention should be paid to situations where key results of the analysis are seriously influenced by a minor perturbation of $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})$. The influence graph of $f_Q(\boldsymbol{\omega})$ is defined as $\gamma(\boldsymbol{\omega}) = (\boldsymbol{\omega}^T, f_Q(\boldsymbol{\omega}))^T$. Since it is difficult to obtain the complete influence graph, the normal curvature $C_{f_Q,h}$ of $\gamma(\boldsymbol{\omega})$ at $\boldsymbol{\omega}^0$ in the direction of a unit vector $\mathbf{h}$ is used to summarize the local behavior of $f_Q(\boldsymbol{\omega})$. Define $\ddot{\mathbf{Q}}_{\omega^0} = \partial^2 Q(\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^T|_{\omega=\omega^0}$, $\ddot{\mathbf{Q}}_\theta(\hat{\boldsymbol{\theta}}) = \partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T|_{\theta=\hat{\theta}}$ and $\boldsymbol{\Delta}_\omega = \partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^T|_{\theta=\theta(\omega)}$. Using a similar derivation as in Cook (1986), it can be shown that the normal curvature $C_{f_Q,h}$ of $\gamma(\boldsymbol{\omega})$ at $\boldsymbol{\omega}^0$ is $C_{f_Q,h} = -2\mathbf{h}^T\ddot{\mathbf{Q}}_{\omega^0}\mathbf{h} = 2\mathbf{h}^T\boldsymbol{\Delta}_{\omega^0}^T\{-\ddot{\mathbf{Q}}_\theta(\hat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{\Delta}_{\omega^0}\mathbf{h}$.

Let $\{(\lambda_i, \mathbf{e}_i), i = 1, \ldots, m\}$ be the eigenvalue-eigenvector pairs of $-2\ddot{\mathbf{Q}}_{\omega^0}$ with $\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_m = 0$ and orthonormal eigenvectors $\{\mathbf{e}_i, i = 1, \ldots, m\}$. The eigenvector $\mathbf{e}_1 = \mathbf{h}_{max}$ corresponding to the largest eigenvalue λ_1 provides important information for judging large enough change of $C_{f_Q,h}$ with respect to a minor perturbation on the postulated model. However, as pointed out recently by Lesaffre and Verbeke (1998), Poon and Poon (1999) and Zhu and Lee (2001), it is not enough to assess local influence by inspecting only $\mathbf{h}_{max}$. Hence, we consider the following aggregated contribution vector of all eigenvectors associated with nonzero eigenvalues. Let $\tilde{\lambda}_i = \lambda_i/\sum_{k=1}^r \lambda_k$, $\mathbf{e}_i^2 = (e_{i1}^2, \ldots, e_{im}^2)^T$, and $M(0) = \sum_{k=1}^r \tilde{\lambda}_k \mathbf{e}_k^2$. The jth component of $M(0)$, $M(0)_j$, is $\sum_{k=1}^r \tilde{\lambda}_k e_{kj}^2$. Assessment of influential cases is based on $\{M(0)_j, j = 1, \ldots, m\}$.

Inspired by Poon and Poon (1999) in modifying Cook's (1986) normal curvature, we define the conformal normal curvature $B_{f_Q,h}$ at $\boldsymbol{\omega}^0$ in a unit direction $\mathbf{h}$ as follows:

$$B_{f_Q,h} = -2\mathbf{h}^T\ddot{\mathbf{Q}}_{\omega^0}\mathbf{h}/tr[-2\ddot{\mathbf{Q}}_{\omega^0}]. \tag{9}$$

Let $\boldsymbol{\omega}_j$ be a basic perturbation vector with jth entry 1 and zero elsewhere. Zhu and Lee (2001) showed that for all j, $M(0)_j = B_{f_Q,\omega_j}$. Note that it is very simple to compute $M(0)_j$, because no eigenvalues and eigenfunctions of $-2\ddot{\mathbf{Q}}_{\omega^0}$ are involved. Let $\bar{M}(0)$ and $SM(0)$ be the mean and standard error of $\{M(0)_j : j = 1, \ldots, m\}$. Clearly, $\bar{M}(0) = 1/m$. It can be used as a benchmark to determine the significance of contribution from an individual case. Similar to Zhu and Lee (2001), $\bar{M}(0)_j + cSM(0)$ may be used as a benchmark, where c is a constant selected on a problem-by-problem basis. In our illustrative example, c is taken to be 1.96. The jth case is regarded as influential if $M(0)_j$ is larger than the benchmark.

The building blocks of the conformal normal curvature involve $\boldsymbol{\Delta}_\omega$ and $\ddot{\mathbf{Q}}_\theta(\boldsymbol{\theta})$ (see (9) and $C_{f_Q,h}$), which are given by: $\ddot{\mathbf{Q}}_\theta(\hat{\boldsymbol{\theta}}) = E[\partial^2 L_c(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T|\mathbf{X}, \mathbf{Z}, \hat{\boldsymbol{\theta}}]$ and $\boldsymbol{\Delta}_{\omega^0} = E[\partial^2 L_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^T|\mathbf{X}, \mathbf{Z}, \hat{\boldsymbol{\theta}}]\ |_{\omega=\omega^0}$. For $h = 1, \ldots, 5$, let $\boldsymbol{\theta}_h$ be the parameter vector that contains the separable unknown parameters in L_h of L_c, and $\ddot{\mathbf{L}}_h(\boldsymbol{\theta}_h) = \partial^2 L_h(\boldsymbol{\theta}_h|\cdot)/\partial\boldsymbol{\theta}_h\partial\boldsymbol{\theta}_h^T$. Then,

$\ddot{\mathbf{L}}_c(\boldsymbol{\theta}) = \partial^2 L_c(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ is a diagonal block matrix with diagonal blocks $\ddot{\mathbf{L}}_1(\boldsymbol{\theta}_1), \ldots, \ddot{\mathbf{L}}_5(\boldsymbol{\theta}_5)$. These derivatives can be obtained by straightforward matrix calculus. Their conditional expectations cannot be evaluated in closed form. So we resort to Monte Carlo integration. Given $\{(\mathbf{F}_w^{(j)}, \mathbf{F}_b^{(j)}, \mathbf{V}^{(j)}), j = 1, \ldots, J\}$, a sufficiently large sample from the conditional distribution $[\mathbf{F}_w, \mathbf{F}_b, \mathbf{V}| \mathbf{X}, \mathbf{Z}, \hat{\boldsymbol{\theta}}]$, the building blocks $\ddot{\mathbf{Q}}_{\omega^0}(\hat{\boldsymbol{\theta}})$ and $\boldsymbol{\Delta}_{\omega^0}$ can be approximated by the sample means of this sample. As this random sample can be obtained from the E-step of the MCEM algorithm in the ML estimation, the required computation is not heavy.

Since the proposed two-level LVM is rather general, there are a number of additional perturbations on top of those commonly considered in local influence analysis. For example, perturbations on the various latent variables (see e.g., perturbation (ii) below) have not been investigated before. Taking the advantage of the fact that the observed-data log-likelihood is decomposed into five separable functions with distinct separable parameters, see (8), it is rather efficient to obtain the diagnostic measures under many different perturbations. Moreover, we see from the definitions of these separable functions that $\{L_1, L_2\}$ and L_3, respectively, relate to the measurement equation and the structural equation of the within-groups model, and that L_4 and L_5, respectively, play the same role in the between-groups model. In applying a perturbation, we know exactly whether it is going to affect the measurement equation or structural equation of the within-groups model or the between-groups model. In general, given a perturbation vector $\boldsymbol{\omega}$, the perturbed complete-data log-likelihood function is

$$\begin{aligned} L_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{X}, \mathbf{Z}, \mathbf{F}_w, \mathbf{F}_b, \mathbf{V}) &= L_{1\omega}(\boldsymbol{\theta}_1, \boldsymbol{\omega}|\mathbf{X}, \mathbf{F}_w, \mathbf{V}) + L_{2\omega}(\boldsymbol{\theta}_2, \boldsymbol{\omega}|\mathbf{Z}, \mathbf{F}_w, \mathbf{V}) \\ &+ L_{3\omega}(\boldsymbol{\theta}_3, \boldsymbol{\omega}|\mathbf{F}_w) + L_{4\omega}(\boldsymbol{\theta}_4, \boldsymbol{\omega}|\mathbf{F}_b, \mathbf{V}) + L_{5\omega}(\boldsymbol{\theta}_5, \boldsymbol{\omega}|\mathbf{F}_b). \end{aligned} \tag{10}$$

Unless otherwise stated, we consider perturbation vectors $\boldsymbol{\omega}$ such that $\boldsymbol{\omega}^0 = (1, \ldots, 1)$. Due to space limitation, only the following perturbations are discussed.

(A) *Local influence relating to the measurement equation of the within-groups model*: To identify an influential observation, a case weights perturbation is usually considered with respect to both the manifest continuous and polytomous entries. However, since the effect of a perturbation on a polytomous data point is very minor, we only need to consider the perturbation on the continuous measurements. For polytomous data, it is more appropriate to identify the influential cells of the corresponding s-dimensional contingency table.

(i) *Case weights perturbation on continuous measurements.* For this perturbation, $L_{h\omega}(\boldsymbol{\theta}_h, \boldsymbol{\omega}|\cdot)$ in $L_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\cdot)$ is equal to $L_h(\boldsymbol{\theta}_h|\cdot)$ for $h = 2, 3, 4, 5$, while

$$L_{1\omega}(\boldsymbol{\theta}_1, \boldsymbol{\omega}|\mathbf{X}, \mathbf{F}_w, \mathbf{V}) = \sum_{g=1}^{G}\sum_{i=1}^{N_g} \omega_{gi} \log p(\mathbf{x}_{gi}|\boldsymbol{\zeta}_{gi}, \mathbf{v}_{g1}, \boldsymbol{\Lambda}_{w1}, \boldsymbol{\Psi}_{w1}).$$

Note that this scheme is a generalization of the case-deletion method.
(ii) *Perturbation on* $\mathbf{\Psi}_{w1}$. In this perturbation, $L_{h\omega}(\boldsymbol{\theta}_h, \boldsymbol{\omega}|\cdot) = L_h(\boldsymbol{\theta}_h|\cdot)$ for $h = 2,3,4,5$, and

$$L_{1\omega}(\boldsymbol{\theta}_1, \boldsymbol{\omega}|\mathbf{X}, \mathbf{F}_w, \mathbf{V}) = \sum_{g=1}^{G}\sum_{i=1}^{N_g} \log p(\mathbf{x}_{gi}|\boldsymbol{\zeta}_{gi}, \mathbf{v}_{g1}, \mathbf{\Lambda}_{w1}, \omega_{gi}^{-1}\mathbf{\Psi}_{w1}).$$

(B) *Local influence relating to the structural equation of the within-groups model*:
(iii) *Perturbation on latent variables* $\boldsymbol{\zeta}_{gi}$. Here, $L_{h\omega}(\boldsymbol{\theta}_h, \boldsymbol{\omega}|\cdot) = L_h(\boldsymbol{\theta}_h|\cdot)$ for $h = 4,5$, and

$$L_{1\omega}(\boldsymbol{\theta}_1, \boldsymbol{\omega}|\mathbf{X}, \mathbf{F}_w, \mathbf{V}) = \sum_{g=1}^{G}\sum_{i=1}^{N_g} \log p(\mathbf{x}_{gi}|\omega_{gi}\boldsymbol{\zeta}_{gi}, \mathbf{v}_{g1}, \mathbf{\Lambda}_{w1}, \mathbf{\Psi}_{w1}),$$

$$L_{2\omega}(\boldsymbol{\theta}_2, \boldsymbol{\omega}|\mathbf{Z}, \mathbf{F}_w, \mathbf{V}) = \sum_{g=1}^{G}\sum_{i=1}^{N_g} \log p(\mathbf{z}_{gi}|\omega_{gi}\boldsymbol{\zeta}_{gi}, \mathbf{v}_{g2}, \mathbf{\Lambda}_{w2}, \mathbf{\Psi}_{w2}, \boldsymbol{\alpha}),$$

$$L_{3\omega}(\boldsymbol{\theta}_3, \boldsymbol{\omega}|\mathbf{F}_w) = \sum_{g=1}^{G}\sum_{i=1}^{N_g} \log p(\omega_{gi}\boldsymbol{\zeta}_{gi}|\mathbf{\Pi}_w, \mathbf{\Gamma}_w, \mathbf{\Phi}_w, \mathbf{\Psi}_{w\delta}).$$

As $\boldsymbol{\zeta}_{gi}$ is closely related to $\mathbf{u}_{gi}$, this perturbation gives some insight about the impact of the analysis with respect to minor change of the latent variable associated with $\mathbf{u}_{gi}$.
(iv) *Perturbation on* $\mathbf{\Psi}_{w\delta}$. In this case, $L_{h\omega}(\boldsymbol{\theta}_h, \boldsymbol{\omega}|\cdot) = L_h(\boldsymbol{\theta}_h|\cdot)$ for $h = 1,2,4,5$, and

$$L_{3\omega}(\boldsymbol{\theta}_3, \boldsymbol{\omega}|\mathbf{F}_w) = \sum_{g=1}^{G}\sum_{i=1}^{N_g} \log p(\boldsymbol{\zeta}_{gi}|\mathbf{\Pi}_w, \mathbf{\Gamma}_w, \mathbf{\Phi}_w, \omega_{gi}^{-1}\mathbf{\Psi}_{w\delta}).$$

(C) *Local influence relating to the between-groups model*: Similar perturbations for assessing local influence on the measurement and structural equations as above may be considered. We just present the following perturbation.
(v) *Case weights perturbation*: In this case, $L_{h\omega}(\boldsymbol{\theta}_h, \boldsymbol{\omega}|\cdot) = L_h(\boldsymbol{\theta}_h|\cdot)$ for $h = 1,2,3,5$, while

$$L_{4\omega}(\boldsymbol{\theta}_4, \boldsymbol{\omega}|\mathbf{F}_b, \mathbf{V}) = \sum_{g=1}^{G} \omega_g \log p(\mathbf{v}_g|\boldsymbol{\zeta}_g, \boldsymbol{\mu}, \mathbf{\Lambda}_b, \mathbf{\Psi}_b).$$

The main computational burden for obtaining the diagnostic measures is the evaluation of $\ddot{\mathbf{Q}}_{\omega^0} = \mathbf{\Delta}_{\omega^0}^T\{-\ddot{\mathbf{Q}}_\theta(\hat{\boldsymbol{\theta}})\}^{-1}\mathbf{\Delta}_{\omega^0}$. In all the perturbations considered, most partitions of $\mathbf{\Delta}_{\omega^0}$ are equal to zero, for example, only $\partial^2 L_{1\omega}(\boldsymbol{\theta}_1, \boldsymbol{\omega}|\cdot)/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\omega}$ is nonzero in perturbation (i). As $\ddot{\mathbf{Q}}(\boldsymbol{\theta})$ is also a diagonal block matrix, the computational burden is not heavy.

5. An Application: Filipino CSWs Study

The illustrative example is based on the study of Morisky et al. (1998) on the effects of establishment policies, knowledge and attitudes on condom use among Filipino commercial sex workers (CSWs). The data set was collected from female CSWs in 97 establishments (bars, night clubs or Karaoke TV) in cities of Philippines. The entire questionnaire consisted of 134 items, covering the areas of attitudes, beliefs, behaviors, self-efficacy for condom use, and social desirability. Nine manifest variables, of which the first three variables are continuous and the remaining are polytomous with a five-point scale, were selected. Questions corresponding to these variables are given in Table 1 of Lee and Shi (2001). For brevity, we deleted those observations with missing entries, and the remaining sample size is 755. The numbers of individuals in establishments varied from 1 to 58, thus this is an unbalanced data set. To unify scales of variables, the raw continuous data were standardized.

Table 1. ML estimates and standard errors of the parameters.

Thresholds			Between-Group			Within-Group		
Par	Est	Std	Par	Est	Std	Par	Est	Std
α_{12}	-1.098	0.050	$\lambda_{b,11}$	0.151	0.035	$\lambda_{w,21}$	1.929	0.226
α_{13}	-0.704	0.055	$\lambda_{b,21}$	0.351	0.055	$\lambda_{w,31}$	1.594	0.084
α_{22}	-0.084	0.039	$\lambda_{b,31}$	0.130	0.029	$\lambda_{w,52}$	0.345	0.087
α_{23}	0.305	0.054	$\lambda_{b,42}$	0.175	0.072	$\lambda_{w,62}$	0.595	0.094
α_{32}	-0.991	0.044	$\lambda_{b,52}$	0.296	0.104	$\lambda_{w,83}$	2.056	0.239
α_{33}	-0.595	0.058	$\lambda_{b,62}$	0.435	0.041	$\lambda_{w,93}$	1.695	0.141
α_{42}	-0.402	0.059	$\lambda_{b,73}$	0.235	0.037	γ_{w1}	-0.051	0.035
α_{43}	0.243	0.043	$\lambda_{b,83}$	0.107	0.014	γ_{w2}	-0.167	0.091
α_{52}	-1.637	0.106	$\lambda_{b,93}$	0.345	0.075	$\psi_{w\delta}$	0.061	0.008
α_{53}	-0.732	0.033						
α_{62}	-1.031	0.074	ψ_{b1}	0.047	0.058	ψ_{w1}	0.869	0.039
α_{63}	-0.119	0.032	ψ_{b2}	0.071	0.036	ψ_{w2}	0.523	0.029
			ψ_{b3}	0.011	0.008	ψ_{w3}	0.800	0.044
			ψ_{b4}	0.063	0.030	ψ_{w3}	0.530	0.061
			ψ_{b5}	0.387	0.088	ψ_{w3}	0.559	0.052
			ψ_{b6}	0.016	0.004	ψ_{w3}	0.698	0.060
			ψ_{b7}	0.027	0.016	ψ_{w3}	0.868	0.065
			ψ_{b8}	0.092	0.014	ψ_{w3}	0.586	0.035
			ψ_{b9}	0.027	0.024	ψ_{w3}	0.661	0.068
			$\phi_{b,21}$	-0.089	0.076	$\phi_{w,11}$	0.371	0.033
			$\phi_{b,31}$	0.208	0.073	$\phi_{w,21}$	0.055	0.018
			$\phi_{b,32}$	-0.095	0.081	$\phi_{w,22}$	0.079	0.006

Three latent variables were used in the measurement equations for models at both levels, using the first three, the second three and the last three manifest variables as indicators for the first, second and third factors, respectively. For the between-groups model, we considered a factor analysis model with the following specifications:

$$\mathbf{\Lambda}_b^T = \begin{bmatrix} \lambda_{b,11} & \lambda_{b,21} & \lambda_{b,31} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & \lambda_{b,42} & \lambda_{b,52} & \lambda_{b,62} & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & \lambda_{b,73} & \lambda_{b,83} & \lambda_{b,93} \end{bmatrix},$$

$$\mathbf{\Phi}_b = \begin{bmatrix} 1.0^* & & \text{sym} \\ \phi_{b,21} & 1.0^* & \\ \phi_{b,31} & \phi_{b,32} & 1.0^* \end{bmatrix}, \text{ and } \mathbf{\Psi}_b = (\psi_{b1}, \psi_{b2}, \psi_{b3}, \psi_{b4}, \psi_{b5}, \psi_{b6}, \psi_{b7}, \psi_{b8}, \psi_{b9}),$$

where any parameter with an asterisk is fixed. For the within-group structure, we consider a LISREL model with a $\mathbf{\Lambda}_w^T$ which has the structure as in $\mathbf{\Lambda}_b^T$ with parameters $\lambda_{w,hk}$,

$$\mathbf{\Phi}_w = \begin{bmatrix} \phi_{w,11} & \phi_{w,12} \\ \phi_{w,21} & \phi_{w,22} \end{bmatrix}, \quad \mathbf{\Psi}_w = (\psi_{w1}, \psi_{w2}, \psi_{w3}, \psi_{w4}, \psi_{w5}, \psi_{w6}, \psi_{w7}, \psi_{w8}, \psi_{w9}),$$

and the following structural equation for the latent variables $\{\eta_{gi}, \xi_{gi1}, \xi_{gi2}\}$: $\eta_{gi} = \gamma_{w1}\xi_{gi1} + \gamma_{w2}\xi_{gi2} + \delta_{gi}$. To identity the polytomous variables, α_{k1} and α_{k4}, $k = 1, \ldots, 6$, were fixed at $\alpha_{kj} = \Phi^{-1}(m_k)$, where m_k is the observed cumulative marginal proportion of the categories with $z_{gik} < j$. There are a total of 54 parameters. ML estimates were obtained by the first MCEM algorithm and local influence was conducted on the basis of the function $Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$. At the rth iteration of the MCEM algorithm, $50+10r$ observations were generated for completing the E-step. Monitoring convergence by bridge sampling, we found that the algorithm converged after about 100 iterations. To be conservative, parameters values at the 200th iteration were taken as the ML estimates. These estimates are reported in Table 1. Hence, we have established a two-level LVM, $\mathbf{u}_{gi} = \mathbf{v}_g + \mathbf{v}_{gi}$, for this data set, where the within-groups model is a LISREL type model with measurement equation $\mathbf{v}_{gi} = \mathbf{\Lambda}_w \boldsymbol{\zeta}_{gi} + \boldsymbol{\epsilon}_{gi}$, and the between-groups model is a confirmatory factor analysis model $\mathbf{v}_g = \mathbf{\Lambda}_b \boldsymbol{\zeta}_g + \boldsymbol{\epsilon}_g$.

The E-step at the last iteration of the MCEM algorithm also produces a sufficiently large sample of the various latent vectors, namely $\{(\mathbf{F}_w^{(j)}, \mathbf{F}_b^{(j)}, \mathbf{V}^{(j)}),\ j = 1, \ldots, J\}$. Hence, we can obtain the following estimates of the latent vectors as by-products: $\hat{\boldsymbol{\zeta}}_{gi} = \sum_{j=1}^J \boldsymbol{\zeta}_{gi}^{(j)}/J$, $\hat{\boldsymbol{\zeta}}_g = \sum_{j=1}^J \boldsymbol{\zeta}_g^{(j)}/J$, $\hat{\mathbf{v}}_g = \sum_{j=1}^J \mathbf{v}_g^{(j)}/J$, where $\boldsymbol{\zeta}_{gi}^{(j)}$, $\boldsymbol{\zeta}_g^{(j)}$ and $\mathbf{v}_g^{(j)}$ are in $\mathbf{F}_w^{(j)}$, $\mathbf{F}_b^{(j)}$ and $\mathbf{V}^{(j)}$, respectively. As a result, for $g = 1, \ldots, G$, $i = 1, \ldots, N_g$, we can get standardized residuals: $\hat{\boldsymbol{\epsilon}}_{gi1}^T \hat{\boldsymbol{\Psi}}_{w1}^{-1} \hat{\boldsymbol{\epsilon}}_{gi1}$, $\hat{\delta}_{gi}^T \hat{\psi}_{w\delta}^{-1} \hat{\delta}_{gi}$, and $\hat{\boldsymbol{\epsilon}}_g^T \hat{\boldsymbol{\Psi}}_b^{-1} \hat{\boldsymbol{\epsilon}}_g$, where $\hat{\boldsymbol{\epsilon}}_{gi1} = \mathbf{x}_{gi} - \hat{\mathbf{v}}_{g1} - \hat{\boldsymbol{\Lambda}}_{w1} \hat{\boldsymbol{\zeta}}_{gi}$, $\hat{\delta}_{gi} = \eta_{gi} - \hat{\gamma}_{w1} \hat{\xi}_{gi1} - \hat{\gamma}_{w2} \hat{\xi}_{gi2}$, and $\hat{\boldsymbol{\epsilon}}_g = \hat{\mathbf{v}}_g - \hat{\boldsymbol{\Lambda}}_b \hat{\boldsymbol{\zeta}}_g$.

We first apply the case weights perturbation on continuous measurements to find influential observations. The mean and standard error of $M(0)_j$ are 0.001 and 0.022, respectively. The benchmark and plots of $M(0)_j$ are presented in Figure 1. In this and other figures, the benchmark is indicated by a dotted horizontal line. The 373th and 554th observations are clearly identified as influential. Hence, if we make a minor perturbation on these individuals, it will strongly influence the outcome of the analysis. We also considered a perturbation such that the $\boldsymbol{\Psi}_{w1}$ of each individual is perturbed to $\omega_{gi}^{-1} \boldsymbol{\Psi}_{w1}$. We found that $\bar{M}(0) = 0.001$, $SM(0) = 0.022$, and index plots of $M(0)_j$ are very similar to those in Figure 1. Again cases 373 and 554 are identified as influential. To identify influential latent variables and study the local influence relating to the structural equation, we considered a perturbation such that the latent variables $\boldsymbol{\zeta}_{gi}$ is perturbed to $\omega_{gi} \boldsymbol{\zeta}_{gi}$. The mean and standard error of $M(0)_j$ are 0.001 and 0.019, respectively. The benchmark and plots of $M(0)_j$ are presented in Figure 2. The 358th, 373th, 433th and 436th latent variables are identified as influential, while the 373th case is most influential. Interestingly, the latent vector corresponding to the 554th manifest observation is not identified as influential.

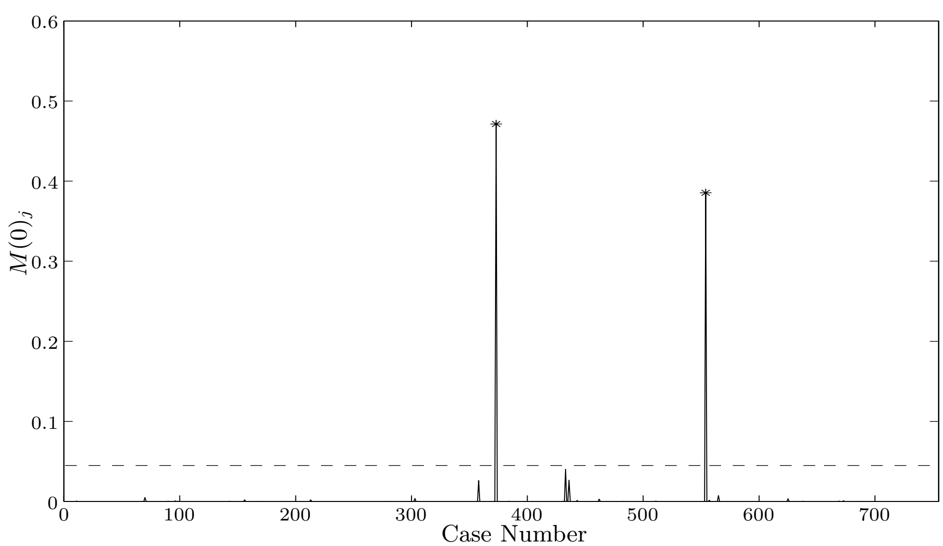

Figure 1. Benchmark and plots of $M(0)_j$ corresponding to case weights perturbation of continuous measurements.

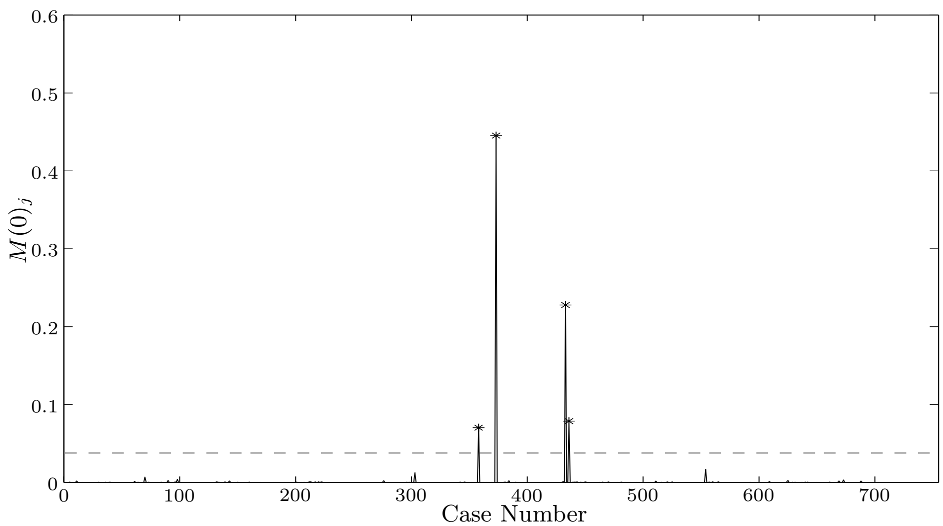

Figure 2. Benchmark and plots of $M(0)_j$ corresponding to perturbation of within-groups latent variables.

To find a possible explanation of the above results, we computed the standardized estimated residuals $\hat{\boldsymbol{\epsilon}}_{gi1}^T \hat{\boldsymbol{\Psi}}_{w1}^{-1} \hat{\boldsymbol{\epsilon}}_{gi1}$ and $\hat{\delta}_{gi}^T \hat{\psi}_{w\delta}^{-1} \hat{\delta}_{gi}$. Plots are presented in Figures 3a and 3b. From Figure 3a, we see that standardized residuals corresponding to the 373th and 554th observations are very large. We examined these observations and found that the variable x_{g3} of the 373th case has an outlying value of 30, while the variable x_{g1} of the 554th case has an outlying value of 99. Comparing to the sample means, these two observations are probably outliers. From Figure 3b, we see that the standardized residuals corresponding to the 358th, 373th, 433th and 436th latent vectors are relatively large. These latent vectors are not fitting the structural equation well. For the 373th case, both its 'measurement equation' and 'structural equation' residuals are large. This may explain why the 373th observation and its associated latent vector are identified as influential. However, for the 554th observation, the 'structural equation' residual of its associated latent vector is relatively small. This may explain why its latent vector not identified as influential.

To illustrate the local influence analysis on the confirmatory factor analysis model associated with the between-groups latent vectors, the plots of $M(0)_j$ and their benchmarks corresponding to the case weights perturbation are displayed in Figure 4. Clearly, the 69th group is most influential in the between-groups model. From plots of $\hat{\boldsymbol{\epsilon}}_g^T \hat{\boldsymbol{\Psi}}_b^{-1} \hat{\boldsymbol{\epsilon}}_g$ (which are not presented to save space), the standardized residual associated with the 69th group is significantly larger than the others. This indicates that the 69th group does not fitting the between-groups model well. By examining the manifest observations in this group, we found that the 554th observation is in there. The analysis of this example is based on a program written in C language. This program is available from the authors upon request.

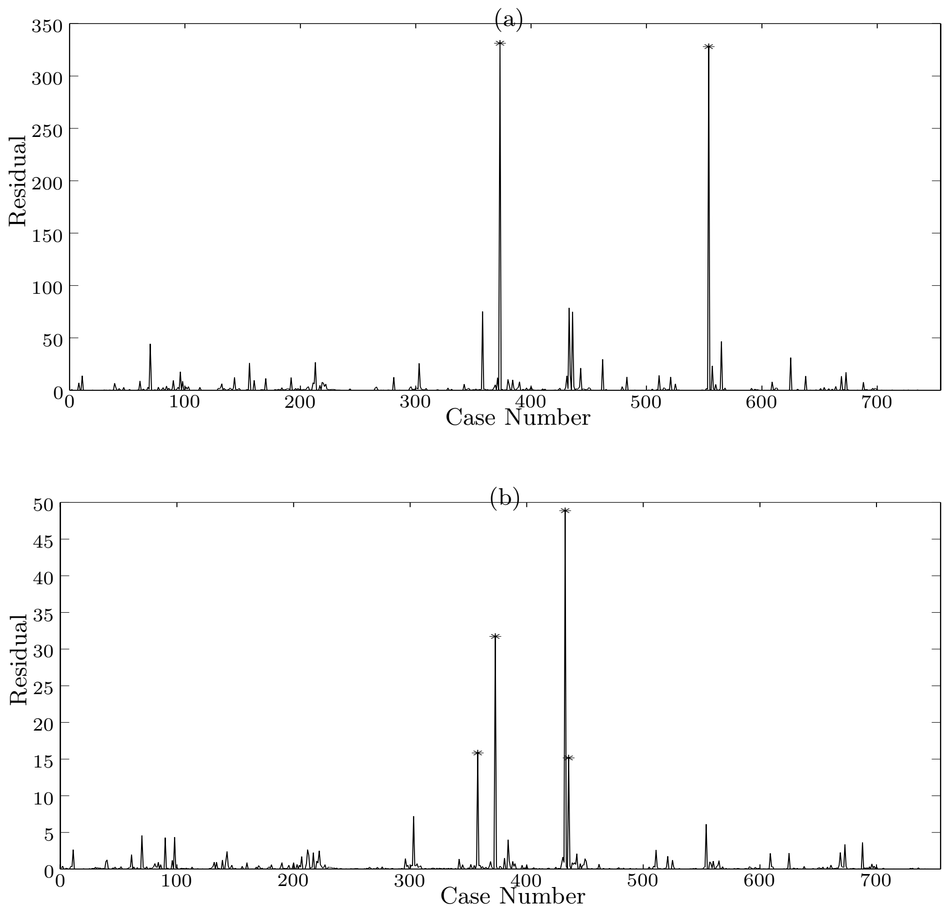

Figure 3. (a) Plots of residuals $\hat{\boldsymbol{\epsilon}}_{gi1}\hat{\boldsymbol{\Psi}}_{w1}^{-1}\hat{\boldsymbol{\epsilon}}_{gi1}$. (b) Plots of residuals $\hat{\delta}_{gi}\hat{\psi}_{w\delta}^{-1}\hat{\delta}_{gi}$.

6. Discussion

Owing to the generality of the proposed model and data structures, our methodology can be applied to assess local influence for many other statistical models as special cases. For example, it can be applied to models in categorical data analysis, and two-level regression models with continuous and polytomous outcomes.

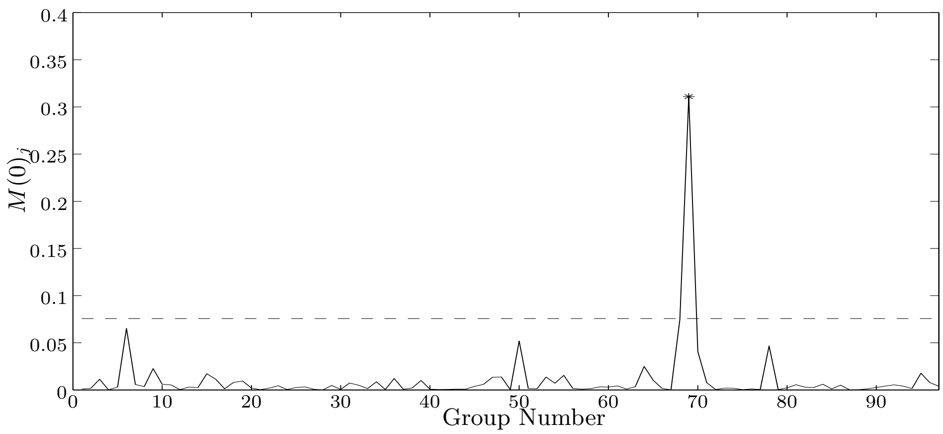

Figure 4. Benchmark and plots of $M(0)_j$ corresponding to case weights perturba-tions of $\mathbf{v}_g$.

Based on the same idea in working with the conditional expectation of the complete-data likelihood in the EM algorithm, global influence analysis of this LVM can be investigated similarly as in Zhu, Lee, Wei and Zhou (2001). Developing such diagnostic measures and comparing them with the local influence measures may be an interesting topic for research.

Acknowledgements

This research is supported by a Hong Kong UGC Earmarked grant, CUHK 4088/99H. We are thankful to D. E. Morisky and J. A. Stein for the use of their AIDS data set, and to the co-editors, an associate editor and a referee for constructive comments.

References

Cook, R. D. (1986). Assessment of local influence (with discussion). *J. Roy. Statist. Soc. Ser. B* **48**, 133-169.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721-741.

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

Lee, S. Y. and Shi, J. Q. (2001). Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics* **57**, 787-794.

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics* **54**, 570-582.

Meng, X, L. and Wong, W. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statist. Sinica* **6**, 831-360.

Morisky, D. E., Tiglao, T. V., Sneed, C. D., Tempongko, S. B., Baltazar, J. C., Detels, R. and Stein, J. A. (1998). The effects of establishment practices, knowledge and attitudes on condom use among Filipino sex workers. *AIDS Care* **10**, 213-220.

Poon, W. Y. and Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *J. Roy. Statist. Soc. Ser. B* **61**, 51-61.

Poon, W. Y., Wang, S. J and Lee, S. Y. (1999). Influence analysis of structural equation models with polytomous variables. *Psychometrika* **64**, 461-473.

Sammel, M. D. and Ryan, L. M. (1996). Latent variables with fixed effects. *Biometrics* **52**, 220-243.

Sammel, M. D., Ryan, L. M. and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *J. Roy. Statist. Soc. Ser. B* **59**, 667-678.

Shi, J. Q. and Lee, S. Y. (2000). Latent variable models with mixed continuous and polytomous data. *J. Roy. Statist. Soc. Ser. B* **62**, 77-87.

Song, X. Y. and Lee, S. Y. (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British J. Math. Statist. Psych.* **54**, 237-263.

Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *J. Amer. Statist. Assoc.* **85**, 699-704.

Zhu, H. T. and Lee, S. Y. (2001). Local influence for incomplete data models. *J. Roy. Statist. Soc. Ser. B* **63**, 111-126.

Zhu, H. T., Lee, S. Y., Wei, B. C. and Zhou, J. (2001). Case deletion measure for models with incomplete data. *Biometrika* **88**, 727-737.

Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

E-mail: xysong@sparc2.sta.cuhk.edu.hk

Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

E-mail: sylee@sparc2.sta.cuhk.edu.hk

(Received December 2001; accepted July 2003)