

SEMIPARAMETRIC LATENT VARIABLE TRANSFORMATION MODELS FOR MULTIPLE MIXED OUTCOMES

Huazhen Lin¹, Ling Zhou¹, Robert M. Elashoff² and Yi Li³

¹*Southwestern University of Finance and Economics,*
²*University of California at Los Angeles and* ³*University of Michigan*

Abstract: Technological advances that allow multiple outcomes to be routinely collected have brought a high demand for valid statistical methods that can summarize and study the latent variables underlying them. Outcome data with continuous and ordinal components present statistical challenges. We develop here a new class of semiparametric latent variable transformation models to summarize the multiple correlated outcomes of mixed types in a data-driven way. We propose a series of estimating equation-based and likelihood-based procedures for estimation and inference. The resulting estimators are shown to be $n^{1/2}$ -consistent (even for nonparametric link functions) and asymptotically normal. Simulations suggest robustness as well as high efficiency, and the proposed approach is applied to assess the effectiveness of recombinant tissue plasminogen activator on ischemic stroke patients.

Key words and phrases: Latent variable model, multiple mixed outcome, normal transformation model, semiparametric.

1. Introduction

Multiple outcomes, measuring diverse aspects of patients' health status, provide more complete and reliable information than traditional single endpoints in clinical studies. Complications arise when observed outcomes consist of components of such mixed types as continuous, binary, and ordinal.

A natural approach, common in the social and biological sciences, is to treat multiple measures as surrogates of an underlying latent variable, and to directly regress the latent variable on the covariates of interest, e.g. treatment. A vast literature has been devoted to continuous multiple outcome data; see O'Brien (1984), Pocock, Geller, and Tsiatis (1987), Legler, Lefkopoulou, and Ryan (1995), Sammel, Lin, and Ryan (1999), Sammel and Ryan (1996), Browne (1984), and Bentler (1983). In contrast, models for mixed-type outcomes are underdeveloped. The related literature has focused primarily on joint models for binary and continuous outcomes in a joint normal framework (Catalano and

Ryan (1992); Cox and Wermuth (1992); Fitzmaurice and Laird (1995); Sammel, Ryan, and Legler (1997); Regan and Catalano (1999); Dunson (2000); Roy and Lin (2000); Gueorguieva and Agresti (2001); Song, Xia, and Lee (2009)), and in a generalized linear model setting (GLLVM, Moustaki (1996); Sammel, Ryan, and Legler (1997); Bartholomew and Knott (1999); Moustaki and Knott (2000); Dunson (2003); Huber, Ronchetti, and Victoria-Feser (2004); Zhu, Eickhoff, and Yan (2005)).

One common theme of existing methods is that the link function relating the observed outcomes to the latent variables is prespecified. For example, the joint normal framework assumes a linear and probit form to combine the continuous and binary outcomes, whereas the generalized linear models typically assume a logit or log function for ordinal outcomes. These parametric assumptions are restrictive and misspecifications can result in improper inference. For example, in our motivating stroke study, ordinal and continuous outcomes are measured, and the traditional joint normal model with a linear link function fails to detect treatment benefit. As elaborated in Section 7, a data-driven link function can establish such benefit.

We develop a semiparametric normal transformation latent variable model to summarize multiple correlated outcomes with continuous and ordinal components. Our method is a flexible yet systematic way of integrating multiple outcomes allowing an unspecified link function. To fix ideas, we consider a case without covariates. As in Muthén (1984), we link the ordinal outcomes to underlying continuous variables. For a continuous variables Y_j with distribution function F_j , $\Phi^{-1}(F_j(Y_j)) \triangleq H_j(Y_j)$ has a standard normal distribution, Φ the normal distribution function. We combine the p -dimensional outcomes Y_1, \dots, Y_p by using functions H_1, \dots, H_p that are data-driven. We propose a series of estimating equation-based and likelihood-based procedures for estimation and inference. Our estimator does not require nonparametric smoothing. We show that the resulting estimators are $n^{1/2}$ -consistent, even for the nonparametric link functions, and asymptotically normal. Finite sample performance of the proposed approach is assessed via simulations, and an application to the effectiveness of recombinant tissue plasminogen activator in a stroke study.

The remainder of the article is organized as follows. The proposed latent variable transformation model is introduced in Section 2. A two-stage estimation procedure is described in Section 3. The asymptotic properties and the variance estimation are considered in Sections 4 and 5, respectively. Simulation results are shown in Section 6, while the analysis results of the ischemic stroke trial are reported in Section 7. We conclude the paper with remarks in Section 8, and defer the technicalities to the Appendix.

2. Models

Suppose there are n randomly selected subjects with p distinct outcomes. For subject $i, i = 1, \dots, n$, we observe the covariate vectors $\mathbf{X}_{i1}, \dots, \mathbf{X}_{ip}$ corresponding to a vector of outcomes $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})^T$. We also observe \mathbf{Z}_i , a vector containing covariates for comparisons, e.g. a treatment indicator. The elements of \mathbf{Y}_i are ordered such that the first p_1 elements are continuous while the remaining $p_2 = p - p_1$ are ordinal. To facilitate joint modeling, we link the ordinal outcomes to underlying continuous variables as in Muthén (1984). Formally, let $Y_{ij} = g_j(Y_{ij}^*; \mathbf{c}_j)$ for $j = 1, \dots, p$, where Y_{ij}^* is a continuous variable underlying Y_{ij} . For the continuous outcomes, we have $Y_{ij} = Y_{ij}^*$, for $j = 1, \dots, p_1$. For the discrete outcomes, with $Y_{ij} \in \{1, \dots, d_j\}$, we have $Y_{ij} = \sum_{l=1}^{d_j} lI(c_{j,l-1} < Y_{ij}^* \leq c_{j,l})$ for $j = p_1 + 1, \dots, p$, where $\mathbf{c}_j = (c_{j,0}, \dots, c_{j,d_j})^T$ are thresholds satisfying $-\infty = c_{j,0} < c_{j,1} < \dots < c_{j,d_j} = \infty$, d_j is the number of categories of the j th outcome. Here, d_j can be arbitrarily large as $n \rightarrow \infty$, so our method can accommodate count data. All of the values of \mathbf{c}_j are unknown. We relate the underlying continuous variables to the latent variable through a semiparametric linear transformation model of the form

$$H_j(Y_{ij}^*) = \mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \mathbf{e}_i + \varepsilon_{ij}, \quad j = 1, \dots, p, \quad (2.1)$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_p^T)^T$ is a vector of regression coefficients, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^T$ are the factor loadings, \mathbf{e}_i is a vector of latent variables summarizing the treatment effect for subject i , and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{ip})^T$ is a vector of independently error distributed as $N(0, \text{diag}(\sigma_1^2, \dots, \sigma_p^2))$. The H_1, \dots, H_p here are unknown increasing transformation functions satisfying $H_j(-\infty) = -\infty$ and $H_j(\infty) = \infty$ for $j = 1, \dots, p$. The last requirement ensures that $\Phi\{a + H_j(-\infty)/b\} = 0$ and $\Phi\{a + H_j(\infty)/b\} = 1$ for any finite a and $b > 0$. If the support of $H_j(\cdot)$ is (a_j, ∞) or $(-\infty, b_j)$, we write $H_j(-\infty) = -\infty$ or $H_j(\infty) = \infty$. This is proper with the monotonicity of H_j .

What distinguishes our model from existing methods lies in nonparametric link functions that are data-driven and do not need to be known a priori. We also remark that, with dummy variables, our method encompasses categorical responses.

We relate a latent variable to \mathbf{Z}_i , which records treatment assignment and other covariates for the sake of comparisons, via

$$\mathbf{e}_i = \boldsymbol{\gamma} \mathbf{Z}_i + \boldsymbol{\epsilon}_i, \quad (2.2)$$

where $\boldsymbol{\gamma}$ is an unknown regression coefficient matrix characterizing the treatment effect in a population, $\boldsymbol{\epsilon}_i$ is the random error distributed as $N(0, \boldsymbol{\Sigma}_e)$, $\boldsymbol{\Sigma}_e =$

$\text{diag}(\sigma_{e_1}^2, \dots, \sigma_{e_q}^2)$, with \mathbf{Z}_i and ϵ_i independent. In general, the number of the latent variables q is less than the number of outcomes p .

Our model is comprehensive and encompasses many well-known models as special cases. To see this, we write $\tilde{\epsilon}_{ij} = \alpha_j e_i + \varepsilon_{ij}$, and rewrite the model for the j -th outcome as

$$H_j(Y_{ij}^*) = \mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \tilde{\epsilon}_{ij}. \quad (2.3)$$

This model belongs to a rich family of semiparametric transformation models. For example, when H_j is a power function, (2.3) reduces to a familiar Box-Cox transformation model (Box and Cox (1964); Bickel and Doksum (1981)). If $H_j(y) = y$ and $H_j(y) = \log(y)$, (2.3) reduces to the additive and multiplicative error models, respectively. More parametric transformation models can be found in Carroll and Ruppert (1988). Han (1987), Cheng, Wei, and Ying (1995), Doksum (1987), Dabrowska and Doksum (1988), Chen, Jin, and Ying (2002), Horowitz (1996), Ye and Duan (1997), Chen (2002), Zhou, Lin, and Johnson (2009), and Lin and Zhou (2009) have proposed regression coefficients and transformation estimators for the model (2.3) with unknown transformation function.

In contrast with the existing semiparametric transformation models, two additional technical difficulties arise for statistical inference based on (2.1) and (2.2): unobserved latent variables e_i are involved, and some outcomes, such as $Y_{ij}^*, j = p_1 + 1, \dots, p$, are not completely observed. We address these issues in the next section.

3. Estimation

Models (2.1) and (2.2) can be rewritten as

$$H_j(Y_{ij}^*) = \mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \boldsymbol{\gamma} \mathbf{Z}_i + \alpha_j^T \epsilon_i + \varepsilon_{ij}, \quad j = 1, \dots, p. \quad (3.1)$$

Hence, given ϵ_i , $H_1(Y_{i1}^*), \dots, H_p(Y_{ip}^*)$ are independent and distributed as $H_j(Y_{ij}^*) \sim N(\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \boldsymbol{\gamma} \mathbf{Z}_i + \alpha_j^T \epsilon_i, \sigma_j^2)$ for $j = 1, \dots, p$. For each given $j > p_1$, the discrete components, we can only estimate $H_j(c_{j,1}), \dots, H_j(c_{j,d_j-1})$, as the \mathbf{c}_j and H_j are unidentifiable separately. To go forward, for each given $j > p_1$, we define a nondecreasing step function G_j with jumps only at $1, \dots, d_j - 1$, and $G_j(m) = H_j(c_{j,m})$ for any $m \in \{1, \dots, d_j - 1\}$, where $c_{j,m}$ is the unknown upper limit of Y_{ij}^* when $Y_{ij} = m$. We also write $G_j = H_j$ for $j \leq p_1$, so the estimation of $H_j, j = 1, \dots, p$, is transferred to the estimation of G_j for $j = 1, \dots, p$.

Equations (3.1) continue to hold if $H_j, \boldsymbol{\beta}_j, \alpha_j$, and σ_j are replaced by $H_j/c, \boldsymbol{\beta}_j/c, \alpha_j/c$ and σ_j/c for any $c > 0$. Here we use $\sigma_j^2 = 1, j = 1, \dots, p$, for scale identification. In addition, we assume that \mathbf{Z}_i and \mathbf{X}_{ij} do not contain intercept term for location normalization. As only $\boldsymbol{\alpha}\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}\boldsymbol{\Sigma}_e\boldsymbol{\alpha}^T$ are identifiable, we take $\sigma_{e,j}^2 = 1$ and $\alpha_{jk} = 0$ for all $j < k$, where $j = 1, \dots, p, k = 1, \dots, q$. Let

$\Theta = \{\beta, \alpha, \gamma\}$ and $\mathbf{G} = \{G_1, \dots, G_p\}$; Θ and \mathbf{G} are the unknown parameters and functions to be estimated in the semiparametric latent variable transformation models (2.1) and (2.2).

3.1. Estimation of Θ

Let $\mathbf{X}_i = \text{diag}(\mathbf{X}_{i1}^T, \dots, \mathbf{X}_{ip}^T)$, $\mathbf{H}_i^{[1]} = (H_1(Y_{i1}^*), \dots, H_{p_1}(Y_{i,p_1}^*))^T$,

$\mathbf{H}_i^{[2]} = (H_{p_1+1}(Y_{i,p_1+1}^*), \dots, H_p(Y_{ip}^*))^T$, and $\mathcal{H}_i^{[2]} = \prod_{j=p_1+1}^p [G_j(Y_{ij} - 1), G_j(Y_{ij})]$.

$\mathbf{H}_i^{[1]}$ is completely observed and $\mathbf{H}_i^{[2]}$ is observed to be belonged to $\mathcal{H}_i^{[2]}$. Since

$$\mathbf{H}_i \equiv (\mathbf{H}_i^{[1]T}, \mathbf{H}_i^{[2]T})^T \sim N(\mathbf{X}_i\beta + \alpha\gamma\mathbf{Z}_i, \Sigma_{22}),$$

where $\Sigma_{22} = \alpha\alpha^T + I_{p \times p}$, the likelihood for the observed data is

$$L(\Theta; \mathbf{G}) \propto |\Sigma_{22}|^{-n/2} \prod_{i=1}^n \int_{\mathbf{x}^{[2]} \in \mathcal{H}_i^{[2]}} \exp \left\{ -\frac{1}{2} \left(\begin{pmatrix} \mathbf{H}_i^{[1]} \\ \mathbf{x}^{[2]} \end{pmatrix} - \mathbf{X}_i\beta - \alpha\gamma\mathbf{Z}_i \right)^T \Sigma_{22}^{-1} \left(\begin{pmatrix} \mathbf{H}_i^{[1]} \\ \mathbf{x}^{[2]} \end{pmatrix} - \mathbf{X}_i\beta - \alpha\gamma\mathbf{Z}_i \right) \right\} d\mathbf{x}^{[2]}. \quad (3.2)$$

The likelihood function involves the infinite dimensional parameter $G_j, j = 1, \dots, p$, so a direct maximization can be prohibitive, especially in the presence of a high dimensional integral. We resort to a two-stage approach. First, we use a series of estimating equations to estimate the transformation functions $G_j, j = 1, \dots, p$. The parameter Θ is then estimated by maximizing a pseudo-likelihood, the likelihood function $L(\Theta; \mathbf{G})$ with \mathbf{G} replaced by its estimated value. We repeat the procedure until convergence.

3.2. Estimation of the transformation function

We first estimate the transformation functions for a given Θ . For any given $j \leq p$, we take $y_j \in \mathcal{R}$ if $j \leq p_1$ and $y_j \in \{1, \dots, d_j\}$ for $j > p_1$, and consider the “marginal” probability for the event of $Y_{ij} \leq y_j$. Then

$$Pr(Y_{ij} \leq y_j | \mathbf{X}_{ij}, \mathbf{Z}_i) = Pr(H_j(Y_{ij}^*) \leq H_j(y_j) | \mathbf{X}_{ij}, \mathbf{Z}_i), \text{ if } j \leq p_1,$$

$$Pr(Y_{ij} \leq y_j | \mathbf{X}_{ij}, \mathbf{Z}_i) = Pr(H_j(Y_{ij}^*) \leq G_j(y_j) | \mathbf{X}_{ij}, \mathbf{Z}_i), \text{ if } j > p_1,$$

both of which are equal to

$$\int_x \Phi(G_j(y_j) - (\mathbf{X}_{ij}^T\beta_j + \alpha_j^T\gamma\mathbf{Z}_i + \alpha_j^T x)) \phi(x) dx,$$

under the convention that $G_j = H_j$ for $j \leq p_1$. Here $\phi(\cdot)$ denotes the density function for a q -dimensional standard normal random vector. This leads to a series of estimating equations

$$\sum_{i=1}^n \left\{ I(Y_{ij} \leq y_j) - \Phi \left(\frac{G_j(y_j) - (\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \boldsymbol{\gamma}_j \mathbf{Z}_i)}{\sqrt{\alpha_j^T \alpha_j + 1}} \right) \right\} = 0, \quad (3.3)$$

for $j = 1, \dots, p$.

Due to the monotonicity of Φ , the estimator $\widehat{G}_j(\cdot)$ of $G_j(\cdot)$ is a nondecreasing step function with jumps only at the observed Y_{ij} , $i = 1, \dots, n$, $j = 1, \dots, p$. Then solving the system of estimating equations (3.3) is equivalent to solving a finite number of equations and, in contrast with traditional nonparametric approaches, our approach does not involve nonparametric smoothing.

For iteratively estimating $\boldsymbol{\Theta}$ and $G_j(\cdot)$, we propose a procedure for choosing initial values. Denote by $\gamma_j = \gamma^T \alpha_j$ for $j = 1, \dots, p$. An application of the double expectation theorem yields

$$E\{\mathbf{X}_{ij} I(Y_{ij} \leq y_j)\} = E\mathbf{X}_{ij} \Phi \left(\frac{G_j(y_j) - (\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \gamma_j^T \mathbf{Z}_i)}{\sqrt{\alpha_j^T \alpha_j + 1}} \right),$$

$$E\{\mathbf{Z}_i I(Y_{ij} \leq y_j)\} = E\mathbf{Z}_i \Phi \left(\frac{G_j(y_j) - (\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \gamma_j^T \mathbf{Z}_i)}{\sqrt{\alpha_j^T \alpha_j + 1}} \right).$$

Let $Y_{(1j)} < \dots < Y_{(d_j, j)}$ be the set of distinct points of Y_{ij} , $i = 1, \dots, n$. Then the initial values of $\boldsymbol{\beta}_j$, γ_j , and $G_j(\cdot)$, $j = 1, \dots, p$, can be obtained by solving the equations

$$\sum_{i=1}^n \left\{ I(Y_{ij} \leq y_j) - \Phi \left(\frac{G_j(y_j) - (\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \gamma_j^T \mathbf{Z}_i)}{\sqrt{\alpha_j^T \alpha_j + 1}} \right) \right\} = 0,$$

for $y_j = Y_{(1j)}, \dots, Y_{(d_j, j)}$,

$$\sum_{i=1}^n \sum_{k=1}^{d_j} \mathbf{X}_{ij} \left\{ I(Y_{ij} \leq Y_{(kj)}) - \Phi \left(\frac{G_j(Y_{(kj)}) - (\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \gamma_j^T \mathbf{Z}_i)}{\sqrt{\alpha_j^T \alpha_j + 1}} \right) \right\} = 0,$$

$$\sum_{i=1}^n \sum_{k=1}^{d_j} \mathbf{Z}_i \left\{ I(Y_{ij} \leq Y_{(kj)}) - \Phi \left(\frac{G_j(Y_{(kj)}) - (\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \gamma_j^T \mathbf{Z}_i)}{\sqrt{\alpha_j^T \alpha_j + 1}} \right) \right\} = 0,$$

for $j = 1, \dots, p$. We set the starting values for α_j , $j = 1, \dots, p$, to be the one satisfying $\alpha_j^T \alpha_j = 1$. The detailed iterative algorithm is provided in Appendix A.

4. Inference in Large Samples

We present the large sample properties of the estimators derived in Section 3. Let $\widehat{\Theta}$ and $\widehat{G}_j, j = 1, \dots, p$, denote the estimators of Θ and $G_j, j = 1, \dots, p$. Throughout, we use the subscript “0” for the true value; for example, G_{j0} is the true value of G_j . Let

$$\mathbf{B} = E \left(\frac{\partial^2 \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta \partial \Theta^T} + \sum_{j=1}^p \frac{\partial^2 \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta \partial G_j(Y_{ij})} d_j^T(Y_{ij}) + \sum_{j=p_1+1}^p \frac{\partial^2 \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta \partial G_j(Y_{ij} - 1)} d_j^T(Y_{ij} - 1) \right),$$

where $L_i(\Theta; \mathbf{G})$ is the contribution of subject i to the likelihood (3.2),

$$d_j(y) = \frac{E \phi \left(\frac{G_{j0}(y) - W_{ij}(\Theta)}{\sqrt{\alpha_j^T \alpha_j + 1}} \right) \left\{ \frac{\partial W_{ij}(\Theta)}{\partial \Theta} + [G_{j0}(y) - W_{ij}(\Theta)] \frac{\partial \log(\sqrt{\alpha_j^T \alpha_j + 1})}{\partial \Theta} \right\}}{E \phi \left(\frac{G_{j0}(y) - W_{ij}(\Theta)}{\sqrt{\alpha_j^T \alpha_j + 1}} \right)} \Bigg|_{\Theta = \Theta_0},$$

and $W_{ij}(\Theta) = \mathbf{X}_{ij}^T \beta_j + \alpha_j^T \gamma \mathbf{Z}_i$.

To facilitate matters, we first assume that \mathbf{B} is negative definite, ensuring the uniqueness of $\widehat{\Theta}$. We assume that the covariates \mathbf{X}_i and \mathbf{Z}_i have bounded supports, and that H is a monotone function.

Theorem 1. *As $n \rightarrow \infty$, $\widehat{\Theta}$ and $\widehat{G}_j(y_j)$ are unique and uniformly consistent for Θ_0 and $G_{j0}(y_j)$ over $y_j \in [a_j, b_j]$ if $j \leq p_1$, and $y_j \in \{1, \dots, d_j - 1\}$ if $j > p_1$.*

Theorem 2. *As $n \rightarrow \infty$, we have*

$$n^{1/2}(\widehat{\Theta} - \Theta_0) \rightarrow N(0, \mathbf{B}^{-1} \mathbf{A} (\mathbf{B}^{-1})^T), \tag{4.1}$$

where \mathbf{A} is defined in Appendix B.

Theorem 3. *As $n \rightarrow \infty$, we have*

$$n^{1/2} \left(\widehat{G}_j(y) - G_{j0}(y) \right) \rightarrow N(0, \Delta_j(y)),$$

for any $y \in [a_j, b_j]$ if $j \leq p_1$, and $y \in \{1, \dots, d_j - 1\}$ if $j > p_1$, where $\Delta_j(y)$ is defined in Appendix B.

We find here that the nonparametric function $G_{j0}(\cdot)$ can be estimated with the parametric convergent rate $n^{-1/2}$. Similar conclusions in different contexts

can be seen in Horowitz (1996), Chen (2002), Ye and Duan (1997) and Zhou, Lin, and Johnson (2009).

5. Estimation of Asymptotic Variance of $\widehat{\Theta}$

As involved computation prohibits the direct usage of the asymptotic variance of $\widehat{\Theta}$, we propose to use a resampling scheme proposed by Jin, Ying, and Wei (2001). Specifically, we first generate n exponential random variables $\xi_i, i = 1, \dots, n$ with mean 1 and variance 1. Fixing the data at their observed values, we solve the following ξ_i -weighted estimation equations and denote the solutions as Θ^* and $G_j^*(y), j = 1, \dots, p$, for any y :

$$\sum_{i=1}^n \xi_i \frac{\partial}{\partial \Theta} \log \left\{ \int_x \prod_{j=1}^{p_1} \phi(G_j(Y_{ij}) - (\mathbf{X}_{ij}^T \beta_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T x)) \right. \\ \times \prod_{j=p_1+1}^p [\Phi(G_j(Y_{ij}) - (\mathbf{X}_{ij}^T \beta_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T x)) \\ \left. - \Phi(G_j(Y_{ij} - 1) - (\mathbf{X}_{ij}^T \beta_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T x))] \phi(x) dx \right\} = 0, \quad (5.1)$$

$$\sum_{i=1}^n \xi_i \left\{ I(Y_{ij} \leq y) - \Phi \left(\frac{G_j(y) - (\mathbf{X}_{ij}^T \beta_j + \alpha_j^T \gamma \mathbf{Z}_i)}{\sqrt{\alpha_j^T \alpha_j + 1}} \right) \right\} = 0, \text{ for } j = 1, \dots, p. \quad (5.2)$$

The estimates Θ^* and $G_j^*(\cdot), j = 1, \dots, p$ can be obtained using the iterative algorithm described in Appendix A. Following Jin, Ying, and Wei (2001), and using the asymptotic expansion (D.4) in Appendix D, we establish the validity of the proposed resampling method.

Proposition. *Under the conditions given in Section 4, the conditional distribution of $n^{1/2}(\Theta^* - \widehat{\Theta})$, given the observed data, converges almost surely to the asymptotic distribution of $n^{1/2}(\widehat{\Theta} - \Theta_0)$.*

Thus we can obtain a large number of realizations of Θ^* , the empirical variance of which can be used to approximate the variance of $\widehat{\Theta}$.

6. Simulation

We investigate the robustness and the efficiency of the proposed method, in comparison with two “extreme” methods. The first method uses (2.1) and (2.2) with misspecified transformation functions, called the MT method; the second uses (2.1) and (2.2) with the correctly specified transformation functions, the CT method. The MT estimator is used to investigate the robustness of the proposed method; the CT estimator is the gold standard by which we evaluate the efficiency of the proposed method. In each case we evaluate the variance estimators

described in Section 5. We assess the performance of the various estimators in terms of bias, standard deviation(SD), and root mean square error(RMSE).

Simulation 1. We simulated 500 datasets, each with 300 subjects. For each subject i , Y_{i1} and Y_{i2} continuous, and Y_{i3} and Y_{i4} , discrete, were generated from the transformation models

$$H_j(Y_{ij}) = \mathbf{X}_i^T \boldsymbol{\beta}_j + \alpha_j e_i + \epsilon_{ij}, \quad j = 1, 2, 3, 4, \quad (6.1)$$

where $H_1(y) = \log(y)$, $H_2(y) = (y^{0.5} - 1)/0.5$, $H_3(y) = y$, $H_4(y) = y^3$. Here Y_{i3}^* and Y_{i4}^* are the underlying continuous variables for Y_{i3} and Y_{i4} , respectively, with $Y_{i3} = \sum_{l=1}^5 lI(c_{l-1,3} < Y_{i3}^* < c_{l,3})$ and $Y_{i4} = \sum_{l=1}^2 (l-1)I(c_{l-1,4} < Y_{i4}^* < c_{l,4})$, where $(c_{0,3}, c_{1,3}, c_{2,3}, c_{3,3}, c_{4,3}, c_{5,3}) = (-\infty, 1, 2, 3, 4, \infty)$ and $(c_{0,4}, c_{1,4}, c_{2,4}, c_{3,4}) = (-\infty, 0, 1, \infty)$. With $\mathbf{X}_i = (X_{1i}, X_{2i})^T$, X_{1i} and X_{2i} were generated independently from the uniform distribution over $[0, 1]$. The regression coefficients were $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})^T = (1.5, 1.5)^T$, $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22})^T = (1, 1)^T$, $\boldsymbol{\beta}_3 = (\beta_{31}, \beta_{32})^T = (2, 2)^T$, and $\boldsymbol{\beta}_4 = (\beta_{41}, \beta_{42})^T = (1, 1)^T$. The loading was $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.5$, and the ϵ_{ij} were independent standard normal. The latent variable e_i was generated from $e_i = Z_i \gamma + \epsilon_i$, where Z_i was uniform on $[0, 1]$, $\gamma = 3$ and ϵ_i standard normal.

Table 1 gives the bias and the standard deviation (SD) of the estimators for the parameters using the proposed method, the CT method, and the MT method with the transformation functions misspecified as $H_1(y) = H_2(y) = H_3(y) = H_4(y) = y$. The results from the MT method were based on 259 replications from the 500 simulation runs, as the Newton-Raphson algorithm failed on 241 replications. Table 1 indicates that the MT estimators have large biases and variances, suggesting that the misspecification of the link function leads to biased and unstable estimates for all the parameters, even for the parameters in the models for the discrete responses where the transformation functions H_3 and H_4 do not matter. In contrast, our method yielded estimates close to the true values, with variances that were very close to those for the CT estimators, suggesting that our procedure is robust with little loss of efficiency.

For each simulated dataset, we also obtained the estimates of the transformations H_1 and H_2 and the threshold parameters. Table 2 presents the average, the standard deviation (SD), and the root of the mean square errors (RMSE) for the threshold parameters. The MT estimator here is severely biased, where our approach yields unbiased estimators with variances close to those of the CT estimators. Figure 1 displays the averaged estimated transformation functions and their 95% empirical pointwise confidence limits based on the 500 simulated datasets; it shows that the proposed estimates of the transformation functions are close to the true transformation functions.

Table 1. Results of the parameter estimation for Simulation 1.

		Proposed	CT	MT			Proposed	CT	MT
β_{11}	Bias	0.019	-0.013	5.405	β_{21}	Bias	0.022	0.001	2.283
	SD	0.190	0.188	3.425		SD	0.195	0.187	0.475
β_{12}	Bias	0.035	0.004	5.662	β_{22}	Bias	0.027	0.005	2.297
	SD	0.186	0.180	3.439		SD	0.198	0.191	0.430
β_{31}	Bias	-0.005	-0.005	-2.328	β_{41}	Bias	0.022	0.027	-12.072
	SD	0.193	0.190	0.773		SD	0.281	0.283	4.800
β_{32}	Bias	0.009	0.007	-2.306	β_{42}	Bias	0.012	0.014	-12.526
	SD	0.196	0.191	0.771		SD	0.286	0.284	10.064
α_1	Bias	-0.003	-0.010	3.893	α_2	Bias	-0.003	-0.011	-0.181
	SD	0.068	0.061	1.530		SD	0.069	0.062	0.084
α_3	Bias	-0.011	-0.007	-0.439	α_4	Bias	-0.003	0.002	0.453
	SD	0.071	0.070	0.052		SD	0.101	0.100	5.792
γ	Bias	0.125	0.095	-1.804					
	SD	0.448	0.408	0.742					

Table 2. The estimates of thresholds for Simulation 1.

		Proposed	CT	MT			Proposed	CT	MT
$G_3(1)$	Bias	-0.004	-0.008	-2.568	$G_3(2)$	Bias	0.005	0.003	-2.832
	SD	0.149	0.139	0.841		SD	0.128	0.119	0.792
$G_3(3)$	Bias	0.008	0.006	-3.110	$G_3(4)$	Bias	0.010	0.009	-3.380
	SD	0.129	0.119	0.752		SD	0.136	0.130	0.709
$G_4(1)$	Bias	-0.002	-0.002	-18.437	$G_4(2)$	Bias	0.021	0.023	-15.122
	SD	0.214	0.211	13.311		SD	0.204	0.201	8.500

Table 3. True and estimated standard errors for Simulation 1.

	SD	SE_{ave}		SD	SE_{ave}
β_{11}	0.190	0.200	β_{12}	0.186	0.220
β_{21}	0.195	0.201	β_{22}	0.198	0.212
β_{31}	0.193	0.193	β_{32}	0.196	0.212
β_{41}	0.281	0.255	β_{42}	0.286	0.250
α_1	0.068	0.074	α_2	0.069	0.083
α_3	0.071	0.071	α_4	0.101	0.090
γ	0.448	0.419			

We also tested the accuracy of the estimation of the standard error given in Section 5. The standard deviations, SD in Tables 1 and 2, based on the 500 simulations, worked well there. To test their accuracy, we took three samples that attained 25%, 50%, and 75% of $ASE = \|\widehat{\Theta} - \Theta_0\|$, respectively, of the 500 simulations. The average of three estimated standard errors based on the 500 realizations of Θ^* , denoted by SE_{ave} , summarizes the overall performance of the standard error estimator, see Table 3.

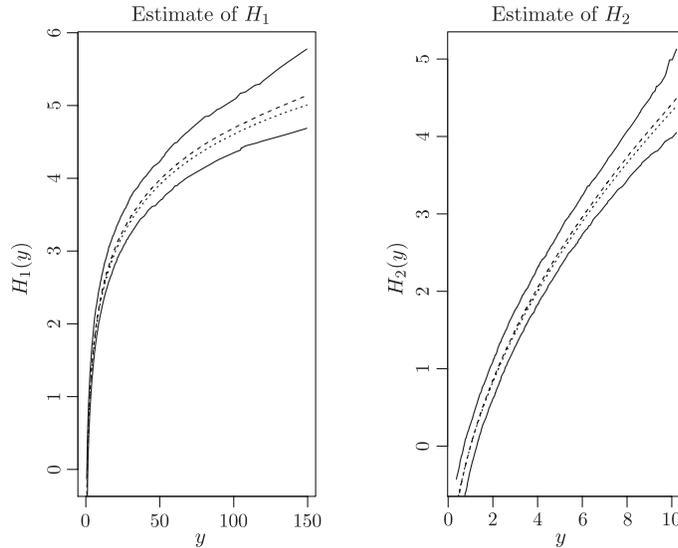


Figure 1. The estimated transformation functions (dotted-lined—true function; solid—95% confidential limit; dashed—average of the estimated transformation function).

Simulation 2 Our method requires Gaussian error. To investigate the sensitivity of our method to this, we generated data according to settings similar to those in Simulation 1, except that we took the outcomes Y_{i1} and Y_{i3} and generated ε_{i1} and ε_{i3} from the centralized and scaled gamma distribution $(Gamma(\tau, 1) - \tau)/\sqrt{\tau}$, which approaches the standard normal as τ increases. We took $\tau = 100, 10, 5, 3,$ and 1. Table 4 presents the bias and SD for the parameters.

The results of the case with $\tau = 1$, marked by *, are based on 418 replications as the algorithm failed to converge in 82 simulations. A useful rule to evaluate the severity of bias, as suggested by Olsen and Schafer (2001), is to check whether the standardized bias (bias over standard deviation) exceeds 0.4. When $\tau \geq 10$, both skewness and excess kurtosis are less than one, the proposed estimators are nearly unbiased. When both skewness and excess kurtosis are around 1 to 2, the proposed estimators are acceptable although they are slightly biased. Only when both the skewness and excess kurtosis are larger than 2 and the error distribution is severely nonnormal, the estimators are biased.

7. Analysis of a Stroke Trial

We analyze an example from a clinical trial to evaluate the effectiveness of an intravenous administration of recombinant tissue plasminogen activator (t-PA) for ischemic stroke (NINDS (1995)). A total of 624 patients were enrolled between January 1991 and October 1994 and were equally randomized to receive either

Table 4. Results of the parameter estimation under different cases for Simulation 2.

		normal	$\tau = 100$	$\tau = 10$	$\tau = 5$	$\tau = 3$	$\tau = 1^*$
Skewness		0	0.2	0.63	0.89	1.15	2
Excess kurtosis		0	0.06	0.6	1.2	2	6
β_{11}	bias	0.013	0.010	0.026	0.038	0.054	0.107
	SD	0.188	0.188	0.188	0.184	0.187	0.206
β_{12}	bias	0.021	0.018	0.027	0.025	0.027	0.111
	SD	0.187	0.188	0.181	0.186	0.193	0.200
β_{31}	bias	0.002	-0.013	-0.009	0.024	0.018	0.074
	SD	0.199	0.206	0.195	0.192	0.196	0.201
β_{32}	bias	0.004	0.013	0.008	0.004	0.012	0.065
	SD	0.206	0.194	0.190	0.192	0.196	0.211
α_1	bias	-0.000	-0.006	-0.019	-0.032	-0.032	-0.050
	SD	0.085	0.090	0.083	0.078	0.084	0.090
α_3	bias	-0.006	-0.014	-0.024	-0.046	-0.043	-0.070
	SD	0.087	0.089	0.086	0.075	0.081	0.084
γ	bias	0.116	0.180	0.243	0.344	0.363	0.652
	SD	0.598	0.636	0.635	0.571	0.650	0.692

t-PA or a placebo. Two primary outcomes the modified Rankin scale (RAN) and NIHSS, were measured three months after the trial began. RAN is a simplified overall assessment of function, a score of 0 indicates the absence of symptoms and a score of 6, severe disability; NIHSS, a measure of neurologic deficit, is on a continuous scale. Baseline blood pressure(BP, X_1), age(X_2), gender(X_3 , 1 = *female*), CT finding Edema indicator (X_4 , 1 = *Edema*), CT finding Mass indicator (X_5 , 1 = *Mass*), weight(X_6), and treatment(Z , 1 = *t-PA*) were included as predictor. The original study (NINDS (1995)) separately compared the difference in each of the outcomes and obtained marginally significant results. Accounting for the intrinsic relationship between RAN and NIHSS, we fit the models

$$\begin{aligned} H_1(Y_1) &= \mathbf{X}^T \beta_1 + \alpha_1 e + \varepsilon_1, \\ H_2(Y_2^*) &= \mathbf{X}^T \beta_2 + \alpha_2 e + \varepsilon_2, \end{aligned} \quad (7.1)$$

where Y_1 is the NIHSS, continuous, and Y_2 is RAN, ordinal. Y_2^* is the underlying continuous variable for Y_2 , with the link $Y_2 = \sum_{l=1}^7 (l-1)I(c_{l-1} < Y_2^* < c_l)$, where $c_0 = -\infty$ and $c_7 = \infty$. $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5, X_6)^T$. The latent variable e is used to evaluate the treatment and is modelled as $e = Z\gamma + \epsilon$.

The resulting estimates of the parameters and standard errors are listed in Tables 5 and 6. The calculation of the standard errors was carried out using the method described in Section 5, based on 1,000 simulations. For comparison purposes, we also applied the traditional joint normal model (JNM) (7.1), with H_1 and H_2 set to be linear functions. For the JNM method, about 50% of the

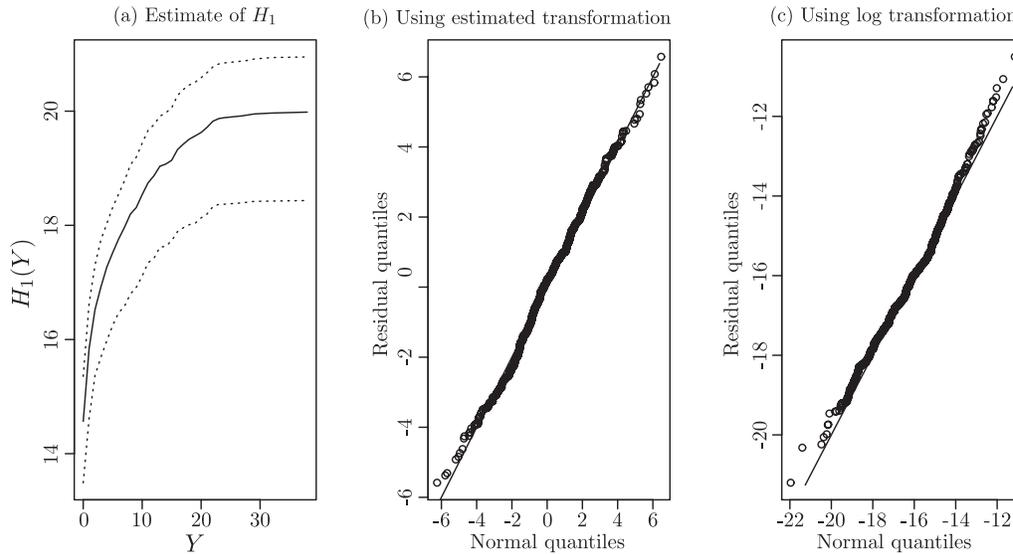


Figure 2. (a) The estimate (Solid) and its 95% confidence limits (dashed) of the transformation function H_1 for the NIHSS; (b) The empirical quantiles of the estimated residuals $\{\hat{\varepsilon}_{i1}\}$ against the normal theoretical quantiles when the transformation functions are estimated by the proposed method for the NIHSS; (c) The empirical quantiles of the estimated residuals $\{\tilde{\varepsilon}_{i2}\}$ against the normal theoretical quantiles when the transformation function is logarithm function for the NIHSS.

runs for the estimation of the variance failed to converge; among the remaining 461 convergent cases, approximately 10% converged to values far away from the estimated parameter values. The standard deviation of the JNM estimator was based on the selected 416 replicates that were the closest to the estimated parameter values over 1,000 replicates. Even with the biased repeated samples that favored the JNM method, our method yielded smaller p-values, suggesting that the proposed method may be more parsimonious in detecting signals. To ascertain the proper transformation function, we display in Figure 2(a) the estimated transformation function and its 95% pointwise confidence limits.

Our analysis revealed that the baseline blood pressure(BP), age, and treatment have significant effects on both the NIHSS and RAN; gender and weight have significant effects on the NIHSS but not on RAN; edema and Mass do not have significant effects on either NIHSS and RAN. The highly significant p-value (0.007) for γ showed that the disease condition is significantly improved after t-PA treatment, where the JNM method failed to detect its benefit, $p=0.093$. Our proposed method confirmed the results that the t-PA treatment is beneficial, as published in the original report.

Table 5. The estimation results of the regression coefficients for the NINDS data using the proposed method and the JNM model. The SDs are based on 1,000 replicates, 416 of which are used to produce the results marked by *.

		Proposed		JNM*	
		β_1	β_2	β_1	β_2
BP	Est.	0.047	0.006	-0.103	-0.016
	p-value	0.000	0.046	0.036	0.347
Age	Est.	0.116	0.029	0.345	0.036
	p-value	0.000	0.000	0.000	0.246
Gender	Est.	0.830	0.183	-2.434	-0.335
	p-value	0.000	0.175	0.252	0.389
Edema	Est.	0.485	0.093	-1.175	0.379
	p-value	0.440	0.854	0.724	0.588
Mass	Est.	0.886	0.920	25.006	3.632
	p-value	0.224	0.089	0.000	0.019
Weight	Est.	0.055	0.006	0.002	-0.005
	p-value	0.000	0.134	0.965	0.739
		α_1	α_2	α_1	α_2
Treat.	Est.	-2.006	-1.247	-9.487	-1.255
	SD	0.135	0.089	0.473	0.552
	p-value	0.000	0.000	0.000	0.023
		γ		γ	
	Est.	0.236		0.407	
	SD	0.087		0.242	
	p-value	0.007		0.093	

Table 6. The estimators of the cutpoints for the NINDS data using the proposed method and the JNM model. The SDs are based on 1,000 replicates, 416 of which are used to produce the results marked by *.

		Proposed			JNM*		
		$G_2(1)$	$G_2(2)$	$G_2(3)$	$G_2(1)$	$G_2(2)$	$G_2(3)$
Est.		1.255	2.369	2.798	-1.828	-0.654	-0.199
SD		0.236	0.217	0.215	3.107	2.875	2.810
		$G_2(4)$	$G_2(5)$	$G_2(6)$	$G_2(4)$	$G_2(5)$	$G_2(6)$
Est.		3.369	4.135	4.502	0.413	1.252	1.668
SD		0.220	0.231	0.244	2.746	2.679	2.652

We checked validity of (7.1) by examining the agreement of the distribution of the estimated residual with that of the normal. Figure 2(b) displays the plot of the empirical quantiles of the estimated residuals, $\{\widehat{\varepsilon}_{i1} = \widehat{H}_1(Y_{i1}) - \mathbf{X}_i^T \widehat{\beta}_1, i = 1, \dots, n\}$, against the normal quantiles. The linearity of the points in Figure 2(b) suggests that the estimated residuals are normally distributed, justifying the assumption of (7.1).

To see whether $\widehat{H}_1(y)$ is the logarithmic function $c \log(y)$, we first obtained $c = 1.27$ by regressing $\widehat{H}_1(Y_{1i})$ on $\log(Y_{1i})$, and computed residuals $\{\tilde{\varepsilon}_{i2} = c \log(Y_{1i}) - X_i^T \widehat{\beta}_1, i = 1, \dots, n\}$. Figure 2(c) displays the empirical quantiles of $\{\tilde{\varepsilon}_{i2}\}$ against the normal theoretical quantiles. The approximate linearity of the points in Figure 2(c) suggests that the estimated transformation $\widehat{H}_1(y)$ is close to a logarithmic function.

8. Discussion

We envision that our method can be extended to accommodate clustered data, such as those arising from repeated measurements in a longitudinal study. Models for multivariate clustered data are complex because they involve two types of correlations: correlation among different outcomes and correlation among repeated measures. We propose to discuss a general methodology for modeling clustered multivariate responses in elsewhere.

Acknowledgements

Lin's research is supported by the National Natural Science Foundation of China (No. 11071197), the National Natural Science Funds for Distinguished Young Scholar of China (No. 11125104), and Program for New Century Excellent Talents in University of China. Li's research is partially supported by grants from the NIH. We thank the Editor, an AE, and two anonymous referees for their insightful suggestions that have significantly improved this manuscript.

Appendix A: Implementation

We outline the algorithm for estimating Θ and $G_j(\cdot), j = 1, \dots, p$ as follows.

Step 0. Choose initial values of the functions $\mathbf{G}^{(0)}(\mathbf{y}) = (G_1^{(0)}(y_1), \dots, G_p^{(0)}(y_p))$ for $\mathbf{y} = \mathbf{Y}_1, \dots, \mathbf{Y}_n$.

Step 1. Given $\mathbf{G}(\mathbf{y})$ at $\mathbf{y} = \mathbf{Y}_1, \dots, \mathbf{Y}_n$, estimate Θ by maximizing (3.2). When $p - p_1$ is large, the computation may be difficult because high-dimensional numerical integration is involved. Note that the dimension of the latent variable in general e_i is low, and rewrite (3.2) as

$$\begin{aligned} & \prod_{i=1}^n \int_x \prod_{j=1}^{p_1} \phi(G_j(Y_{ij}) - (\mathbf{X}_{ij}^T \beta_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T x)) \\ & \quad \times \prod_{j=p_1+1}^p [\Phi(G_j(Y_{ij}) - (\mathbf{X}_{ij}^T \beta_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T x)) \\ & \quad - \Phi(G_j(Y_{ij} - 1) - (\mathbf{X}_{ij}^T \beta_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T x))] \phi(x) dx, \quad (\text{A.1}) \end{aligned}$$

which is a low-dimensional integration. Replacing the integral with the sampling mean, estimate Θ by maximizing the likelihood

$$\prod_{i=1}^n \sum_{k=1}^R \left\{ \prod_{j=1}^{p_1} \phi(G_j(Y_{ij}) - (\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \boldsymbol{\gamma} \mathbf{Z}_i + \alpha_j^T y_k)) \right. \\ \times \prod_{j=p_1+1}^p [\Phi(G_j(Y_{ij}) - (\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \boldsymbol{\gamma} \mathbf{Z}_i + \alpha_j^T y_k)) \\ \left. - \Phi(G_j(Y_{ij} - 1) - (\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \boldsymbol{\gamma} \mathbf{Z}_i + \alpha_j^T y_k))] \right\}, \quad (\text{A.2})$$

where y_1, \dots, y_R are independent standard normal random variables.

Step 2. Given Θ , estimate $\mathbf{G}(\mathbf{y})$ at $\mathbf{y} = \mathbf{Y}_1, \dots, \mathbf{Y}_n$ using (3.3).

Step 3. Repeat Steps 1 and Step 2 until convergence.

Step 4. For every \mathbf{y} in the range of \mathbf{Y} , the estimates of $\mathbf{G}(\mathbf{y})$, $\widehat{\mathbf{G}}(\mathbf{y})$, are obtained by solving (3.3) for $G_j(y_j)$, $j = 1, \dots, p$, by replacing Θ with its estimator from the iteration described here.

Appendix B: Notation

Let $\tilde{\sigma}_j = \sqrt{\alpha_j^T \alpha_j + 1}$, $W_{ij}(\Theta) = \mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \boldsymbol{\gamma} \mathbf{Z}_i$,

$$\psi(y_j) = E \phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta_0)}{\tilde{\sigma}_{j0}} \right),$$

$$\xi_{ij}(y) = I(Y_{ij} \leq y) - \Phi \left(\frac{G_{j0}(y) - W_{ij}(\Theta_0)}{\tilde{\sigma}_{j0}} \right),$$

$$\varphi_{kj1} = E \left\{ \frac{\partial^2 \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta \partial G_j(Y_{ij})} \frac{\tilde{\sigma}_{j0}}{\psi(Y_{ij})} \xi_{kj}(Y_{ij}) | \mathbf{Y}_k, \mathbf{X}_k, \mathbf{Z}_k \right\},$$

$$\varphi_{kj2} = E \left\{ \frac{\partial^2 \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta \partial G_j(Y_{ij} - 1)} \frac{\tilde{\sigma}_{j0}}{\psi(Y_{ij} - 1)} \xi_{kj}(Y_{ij} - 1) | \mathbf{Y}_k, \mathbf{X}_k, \mathbf{Z}_k \right\}.$$

Let $\varpi_i = \frac{\partial \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta} + \sum_{j=1}^p \varphi_{ij1} + \sum_{j=p_1+1}^p \varphi_{ij2}$, $\mathbf{A} = \mathbf{E}(\varpi_i^{\otimes 2})$.

Take

$$\Delta_j(y) = \frac{\alpha_{j0}^T \alpha_{j0} + 1}{\psi^2(y)} E \{ \xi_{ij}(y) + \mathbf{D}^T(y) \mathbf{B}^{-1} \varpi_i \}^2,$$

$$\mathbf{D}(y) = E \phi \left(\frac{G_{j0}(y) - W_{ij}(\Theta)}{\tilde{\sigma}_j} \right) \left\{ [G_{j0}(y) - W_{ij}(\Theta)] \frac{\partial \tilde{\sigma}_j^{-1}}{\partial \Theta} - \frac{\partial W_{ij}(\Theta)}{\tilde{\sigma}_j \partial \Theta} \right\} \Bigg|_{\Theta = \Theta_0}.$$

Appendix C: Proof of Theorem 1

It follows from the Uniform Law of Large Numbers and the monotonicity of H_0 that for any $\eta \geq 0, \zeta > 0$, uniformly in $y_j \in \mathcal{R} \equiv (-\infty, \infty), j = 1, \dots, p$, and $\Theta \in D_\eta = \{\Theta : \|\Theta - \Theta_0\| \leq \eta\}$,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \left\{ I(Y_{ij} \leq y_j) - \Phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta)}{\tilde{\sigma}_j} - \zeta \right) \right\} \\ & \rightarrow E \left\{ \Phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta_0)}{\tilde{\sigma}_{j0}} \right) - \Phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta)}{\tilde{\sigma}_j} - \zeta \right) \right\}, \end{aligned} \tag{C.1}$$

almost surely as $n \rightarrow \infty$, where $W_{ij}(\Theta) = \mathbf{X}_{ij}^T \beta_j + \alpha_j^T \gamma \mathbf{Z}_i$. Uniform convergence follows from empirical process techniques. Indeed, as $\frac{G_{j0}(y_j) - W_{ij}(\Theta)}{\tilde{\sigma}_j}$ can be regarded as a linear function class on \mathcal{R}^d and is thus VC, by the monotonicity of Φ , $\Phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta)}{\tilde{\sigma}_j} - \zeta \right)$ is also VC. Moreover, as the indicator function class is VC and both the indicator function and $\Phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta)}{\tilde{\sigma}_j} - \zeta \right)$ are bounded by 1, the uniform convergence of (C.1) follows from Van de Geer (2000).

It follows from (C.1) that for large $n, y_j \in \mathcal{R}, \Theta \in D_\eta$, and sufficiently large ζ ,

$$\frac{1}{n} \sum_{i=1}^n \left\{ I(Y_{ij} \leq y_j) - \Phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta)}{\tilde{\sigma}_j} - \zeta \right) \right\} > 0, \tag{C.2}$$

$$\frac{1}{n} \sum_{i=1}^n \left\{ I(Y_{ij} \leq y_j) - \Phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta)}{\tilde{\sigma}_j} + \zeta \right) \right\} < 0. \tag{C.3}$$

This together with the monotonicity and continuity of Φ implies that there exists a unique $\widehat{G}_j(y_j; \Theta)$ such that

$$\frac{1}{n} \sum_{i=1}^n \left\{ I(Y_{ij} \leq y_j) - \Phi \left(\frac{\widehat{G}_j(y_j; \Theta) - W_{ij}(\Theta)}{\tilde{\sigma}_j} \right) \right\} = 0. \tag{C.4}$$

By differentiating both side of (C.4) with respect to Θ , we obtain the identity

$$\begin{aligned} & \frac{\partial \widehat{G}_j(y_j; \Theta)}{\partial \Theta} \\ & = \frac{\sum_{i=1}^n \phi \left(\frac{\widehat{G}_j(y_j; \Theta) - W_{ij}(\Theta)}{\tilde{\sigma}_j} \right) \left\{ \frac{\partial W_{ij}(\Theta)}{\partial \Theta} + \left[\widehat{G}_j(y_j; \Theta) - W_{ij}(\Theta) \right] \frac{\partial \log \tilde{\sigma}_j}{\partial \Theta} \right\}}{\sum_{i=1}^n \phi \left(\frac{\widehat{G}_j(y_j; \Theta) - W_{ij}(\Theta)}{\tilde{\sigma}_j} \right)}. \end{aligned} \tag{C.5}$$

When $\Theta = \Theta_0$, (C.2) and (C.3) hold for any $\zeta > 0$, and we have that $\widehat{G}_j(y_j; \Theta_0)$

→ $G_0(y_j)$ uniformly in $y_j \in \mathcal{R}$. Hence

$$\frac{\partial \widehat{G}_j(y_j; \Theta_0)}{\partial \Theta} \rightarrow \frac{E\phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta)}{\bar{\sigma}_j} \right) \left\{ \frac{\partial W_{ij}(\Theta)}{\partial \Theta} + [G_{j0}(y_j) - W_{ij}(\Theta)] \frac{\partial \log \bar{\sigma}_j}{\partial \Theta} \right\}}{E\phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta)}{\bar{\sigma}_j} \right)} \Bigg|_{\Theta = \Theta_0} \equiv d_j(y_j). \tag{C.6}$$

To show the existence and uniqueness of $\widehat{\Theta}$, we let $W(\Theta; \mathbf{G}) = \frac{\partial \log L(\Theta; \mathbf{G})}{\partial \Theta}$ and $S(\Theta) = W(\Theta; \widehat{\mathbf{G}}(\Theta))/n$, which is $W(\Theta; \mathbf{G})$ with $G_j(\cdot), j = 1, \dots, p$, replaced by $\widehat{G}_j(\cdot; \Theta), j = 1, \dots, p$. It follows from (C.5), and $\widehat{G}_j(y_j; \Theta_0) \rightarrow G_{j0}(y_j)$ uniformly in $y_j \in \mathcal{R}$, that

$$\frac{\partial S(\Theta_0)}{\partial \Theta^T} = \frac{1}{n} \left\{ \frac{\partial W(\Theta; \mathbf{G})}{\partial \Theta^T} + \sum_{i=1}^n \left(\sum_{j=1}^p \frac{\partial^2 \log L_i(\Theta_0; \mathbf{G})}{\partial \Theta \partial G_j(Y_{ij})} \frac{\partial \widehat{G}_j(Y_{ij}; \Theta)}{\partial \Theta^T} + \sum_{j=p_1+1}^p \frac{\partial^2 \log L_i(\Theta_0; \mathbf{G})}{\partial \Theta \partial G_j(Y_{ij} - 1)} \frac{\partial \widehat{G}_j(Y_{ij} - 1; \Theta)}{\partial \Theta^T} \right) \right\} \Bigg|_{\mathbf{G} = \widehat{\mathbf{G}}(\Theta), \Theta = \Theta_0} \rightarrow \mathbf{B},$$

where \mathbf{B} is defined in Section 4. Now, because $S(\Theta_0) \rightarrow 0$ and \mathbf{B} is negative definite, there exists a unique solution $\widehat{\Theta}$ to the equation $S(\Theta) = 0$ in a neighborhood of Θ_0 . The foregoing proof also implies that $\widehat{\Theta}$ is strongly consistent and that $\widehat{G}_j(y_j) = \widehat{G}_j(y_j; \widehat{\Theta}) \rightarrow G_{j0}(y_j)$ almost surely uniformly in $y_j \in \mathcal{R}$. Thus Theorem 1 follows.

Appendix D: Proof of Theorem 2

By the consistency of $\widehat{\Theta}$ and a Taylor series expansion of $S(\widehat{\Theta})$ around Θ_0 , we get

$$\widehat{\Theta} - \Theta_0 \approx -\mathbf{B}^{-1}S(\Theta_0). \tag{D.1}$$

Note that

$$\begin{aligned} S(\Theta_0) &= \left\{ n^{-1} \frac{\partial \log L(\Theta_0; \mathbf{G}_0)}{\partial \Theta} + n^{-1} \frac{\partial \log L(\Theta_0; \widehat{\mathbf{G}}(\Theta_0))}{\partial \Theta} - n^{-1} \frac{\partial \log L(\Theta_0; \mathbf{G}_0)}{\partial \Theta} \right\} \\ &\approx n^{-1} \frac{\partial \log L(\Theta_0; \mathbf{G}_0)}{\partial \Theta} \\ &\quad + n^{-1} \sum_{i=1}^n \left\{ \sum_{j=1}^p \frac{\partial \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta \partial G_j(Y_{ij})} \left(\widehat{G}_j(Y_{ij}; \Theta_0) - G_{j0}(Y_{ij}) \right) \right. \\ &\quad \left. + \sum_{j=p_1+1}^p \frac{\partial \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta \partial G_j(Y_{ij} - 1)} \left(\widehat{G}_j(Y_{ij} - 1; \Theta_0) - G_{j0}(Y_{ij} - 1) \right) \right\}. \tag{D.2} \end{aligned}$$

Because of (C.4), we have

$$\begin{aligned} \widehat{G}_j(y_j; \Theta_0) - G_{j0}(y_j) &= \frac{\tilde{\sigma}_{j0}}{n\psi(y_j)} \sum_{i=1}^n \left\{ I(Y_{ij} \leq y_j) - \Phi \left(\frac{G_{j0}(y_j) - W_{ij}(\Theta_0)}{\tilde{\sigma}_{j0}} \right) \right\} \\ &\quad + o_p(n^{-1/2}), \end{aligned} \tag{D.3}$$

where $\psi(y_j)$ is defined in Section 4. Substituting (D.3) into (D.2) and exchanging the summations, we get

$$S(\Theta_0) \approx n^{-1} \frac{\partial \log L(\Theta_0; \mathbf{G}_0)}{\partial \Theta} + n^{-1} \sum_{i=1}^n \left\{ \sum_{j=1}^p \varphi_{ij1} + \sum_{j=p_1+1}^p \varphi_{ij2} \right\}.$$

Hence, by (D.1), we have

$$\widehat{\Theta} - \Theta_0 \approx -n^{-1} \mathbf{B}^{-1} \sum_{i=1}^n \left\{ \frac{\partial \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta} + \sum_{j=1}^p \varphi_{ij1} + \sum_{j=p_1+1}^p \varphi_{ij2} \right\}. \tag{D.4}$$

The proof of Theorem 2 is completed.

Appendix E: Proof of Theorem 3

Because of (C.4), for any $y \in \mathcal{R}$ we have

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \left\{ I(Y_{ij} \leq y) - \Phi \left(\frac{G_{j0}(y) - W_{ij}(\Theta_0)}{\tilde{\sigma}_{j0}} \right) \right\} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left\{ \Phi \left(\frac{G_{j0}(y) - W_{ij}(\Theta_0)}{\tilde{\sigma}_{j0}} \right) - \Phi \left(\frac{G_{j0}(y) - W_{ij}(\widehat{\Theta})}{\sqrt{\widehat{\alpha}_j^T \alpha_j + 1}} \right) \right\} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left\{ \Phi \left(\frac{G_{j0}(y) - W_{ij}(\widehat{\Theta})}{\sqrt{\widehat{\alpha}_j^T \alpha_j + 1}} \right) - \Phi \left(\frac{\widehat{G}_j(y) - W_{ij}(\widehat{\Theta})}{\sqrt{\widehat{\alpha}_j^T \alpha_j + 1}} \right) \right\} = 0, \end{aligned}$$

hence

$$\frac{1}{n} \sum_{i=1}^n \xi_{ij}(y) - \mathbf{D}^T(y) (\widehat{\Theta} - \Theta_0) - \frac{\psi(y)}{\tilde{\sigma}_{j0}} (\widehat{G}_j(y) - G_{j0}(y)) = o_p(n^{-1/2}),$$

where $\mathbf{D}(y)$ is defined in Appendix B. Substituting (D.4) into the equation above we obtain

$$\begin{aligned} &\widehat{G}_j(y) - G_{j0}(y) \\ &= \frac{\tilde{\sigma}_{j0}}{n\psi(y)} \sum_{i=1}^n \left\{ \xi_{ij}(y) + \mathbf{D}^T(y) \mathbf{B}^{-1} \left(\frac{\partial \log L_i(\Theta_0; \mathbf{G}_0)}{\partial \Theta} + \sum_{j=1}^p \varphi_{ij1} + \sum_{j=p_1+1}^p \varphi_{ij2} \right) \right\} \\ &\quad + o_p(n^{-1/2}). \end{aligned} \tag{E.1}$$

The proof of Theorem 3 is completed.

References

- Bartholomew, D. J. and Knott, M. (1999). *Latent Variable Models and Factor Analysis*. Arnold, London.
- Bentler, P. M. (1983). Some contributions to efficient statistics for structural models: Specification and estimation of moment structures. *Psychometrika* **48**, 493-517.
- Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.* **76**, 296-311.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. Ser. B B* **26**, 211-252.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British J. Math. Statist. Psych.* **37**, 62-83.
- Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall.
- Catalano, P. J. and Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *J. Amer. Statist. Assoc.* **87**, 651-658.
- Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659-668.
- Chen, S. (2002). Rank estimation of transformation models. *Econometrica* **70**, 1683-1697.
- Cheng, S. C., Wei, L. J., and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835-845.
- Cox, D. R. and Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika* **79**, 441-461.
- Dabrowska, D. M. and Doksum, K. A. (1988). Partial likelihood in transformation models with censored data. *Scand. J. Statist.* **15**, 1-23.
- Doksum, K. A. (1987). An extension of partial likelihood methods for proportional hazard models to general transformation models. *Ann. Statist.* **15**, 325-345.
- Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *J. Roy. Statist. Soc. Ser. B* **62**, 355-366.
- Dunson, D. B. (2003). Dynamic latent trait models for multidimensional longitudinal data. *J. Amer. Statist. Assoc.* **98**, 555-563.
- Fitzmaurice, G. M. and Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *J. Amer. Statist. Assoc.* **90**, 845-852.
- Gueorguieva, R. V. and Agresti, A. (2001). A correlated probit model for joint modelling of clustered binary and continuous responses. *J. Amer. Statist. Assoc.* **96**, 1102-1112.
- Han, A. K. (1987). Non-parametric analysis of a generalized regression model, *J. Econom.* **35**, 303-316.
- Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variables. *Econometrica* **64**, 103-137.
- Huber, P., Ronchetti, E., and Victoria-Feser, M. (2004). Estimation of generalized linear latent variable models. *J. Roy. Statist. Soc. Ser. B* **66**, 893-908.
- Jin, Z., Ying, Z. and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381-390.
- Legler, J. M., Lefkopoulou, M., and Ryan, L. M. (1995). Efficiency and power of tests for multiple binary outcomes. *J. Amer. Statist. Assoc.* **90**, 680-693.

- Lin, H. and Zhou, X. H. (2009). A semi-parametric two-part mixed-effects heteroscedastic transformation model for correlated right-skewed semi-continuous data. *Biostatistics* **10**, 640-658.
- Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British J. Math. Statist. Psych.* **49**, 313-334.
- Moustaki, I. and Knott, M. (2000). Generalized latent trait models. *Psychometrika* **65**, 391-411.
- Muthén, D. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* **49**, 115-132.
- National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group (NINDS) (1995). Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med.* **333**, 1581-1587.
- O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* **40**, 1079-1087.
- Olsen, M. K. and Schafer, J. (2001). A two-part random-effects model for semi-continuous longitudinal data. *J. Amer. Statist. Assoc.* **96**, 730-745.
- Pocock, S. T., Geller, N. L., and Tsatis, A. A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics* **43**, 487-498.
- Regan, M. M., and Catalano, P. J. (1999). Likelihood models for clustered binary and continuous outcomes: application to developmental toxicology. *Biometrics* **55**, 760-768.
- Roy, J., and Lin, X. (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* **56**, 1047-1054.
- Sammel, M. D., Lin, X., and Ryan, L. (1999). Multivariate linear mixed models for multiple outcomes. *Statist. Medicine* **18**, 2479-2492.
- Sammel, M. D. and Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics* **52**, 650-663.
- 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.
- Song, X. Y., Xia, Y. M., and Lee, S. Y. (2009). Bayesian semiparametric analysis of structural equation models with mixed continuous and unordered categorical variables. *Statistics in Medicine* **28**, 2253-2276.
- Van de Geer, S. (2000). *Empirical Processes in M-estimation*. Cambridge Univ. Press.
- Ye, J. M. and Duan, N. H. (1997). Nonparametric $n^{-1/2}$ -consistent estimation for the general transformation models. *Ann. Statist.* **25**, 2682-2717.
- Zhou, X. H., Lin, H., and Johnson, E. (2009). Nonparametric heteroscedastic transformation regression models for skewed data with an application to health care costs. *J. Roy. Statist. Soc. Ser. B* **70**, 1029-1047.
- Zhu, J., Eickhoff, J. C., and Yan, P. (2005). Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* **61**, 674-683.

Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China.

E-mail: linhz@swufe.edu.cn

Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China.

E-mail: zhouling1003@126.com

Department of Biomathematics, University of California at Los Angeles, Los Angeles CA 90095-1772, USA.

E-mail: relashof@biomath.ucla.edu

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, USA.

E-mail: yili@umich.edu

(Received November 2012; accepted April 2013)