Semiparametric latent variable transformation models for multiple mixed outcomes

Huazhen Lin*, Ling Zhou†, Robert M. Elashoff ‡, Yi Li§

SUMMARY The surge of technological advances that allow multiple outcomes to be routinely collected has brought up a high demand of valid statistical methods that can summarize and study the latent variables underlying them. Mixed outcome data, e.g. those with continuous and ordinal components, present further statistical challenges. Addressing to these challenges, we develop 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 the nonparametric link functions) and asymptotically normal. Simulations suggest robustnees 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: Latent variable model, multiple mixed outcome, normal transformation model, semiparametric.

*Address for correspondence: School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China,611130. *linhz@swufe.edu.cn*

†School of Statistics, Southwestern University of Finance and Economics. *zhouling1003@126.com*

‡Department of Biomathematics, University of California, Los Angeles. *relashof@biomath.ucla.edu*

§Department of Biostatistics, University of Michigan, USA. *yili@umich.edu*

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 as in many situations the observed outcomes consist of components of mixed types, e.g. continuous, binary and ordinal. It is of substantial interest to study how to combine the mixtures of these continuous and discrete data to obtain prognostic factors for patients' health status.

A natural approach, as commonly used in 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 amount of literature has been devoted to continous multiple outcome data; see OBrien (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. For example, 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 *et al.*, 1997; Regan and Catalano, 1999; Dunson, 2000; Roy and Lin, 2000; and Gueorguieva and Agresti, 2001; Song *et al.*, 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 et al., 2004; Zhu, Eickhoff and Yan, 2005).

One common theme of these existing methods is that the link function relating the observed outcomes to the latent variables has to be prespecified. That is, these methods combined the multiple outcomes in a prespecified form. 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. However, these parametric assumptions on the link functions tend to be rather restrictive and the misspecifications can result in improper or wrong inference for the mixtures of continuous and ordinal responses. The link selection is crucial in that the validity of the fitted model as well as its inference heavily depends on whether the link function is specified correctly. For example, in our motivating stroke study, two types of outcomes, ordinal and continuous, are measured. The traditional joint normal model with a linear link function fails to detect the benefit of treatment. On the hand, as elaborated in Section 7, a data-driven link function successfully established such benefit.

In the paper, we develop a semiparametric normal transformation latent variable model to summarize the multiple correlated outcomes with continuous and ordinal components. Our method is a flexible yet systematic way of integrating multiple outcomes by allowing the link function unspecified. To fix the idea, we consider a case without covariates. As in Muthén (1984), we first link the ordinal outcomes to some underlying continuous variables. Then for a continuous variables Y_j with a distribution function F_j, its probit-type transformation $\Phi^{-1}(F_j(Y_j))\widehat{\equiv}H_j(Y_j)$ follows a standard normal distribution, where Φ is the standard normal distribution. Since the latent variables are normal, it is natural to impose a linear form connecting the normal random fields $H_j(Y_j)$ and the normal latent variables. That is, we combine the $p-$dimensional outcomes $Y_1, \ldots, Y_p$ by using functions $H_1, \ldots, H_p$, which are all 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 and, hence avoids complicated smoothing-related problems including selection of smoothing parameters. 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 in assessing the effectiveness of recombinant tissue plasminogen activator in the aforementioned 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 derived in Sections 4 and 5, respectively. Simulation results are shown in Section 6, while the analysis results of the ischemic stroke trial is reported in Section 7. We conclude the paper with concluding remarks in Section 8 and defer all the technical proofs and notations to the Appendix.

2 Models

Suppose there are n randomly selected subjects with p distinct outcomes. For subject $i, i = 1, \ldots, n$, we observe the covariate vectors $\mathbf{X}_{i1}, \cdots, \mathbf{X}_{ip}$ corresponding to a vector of outcomes $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{ip})^T$. We also observe $\mathbf{Z}_i$, a vector containing covariates for comparisons, e.g. 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 the underlying continuous variables as in Muthén (1984). Formally, let $Y_{ij} = g_j(Y_{ij}^*; \mathbf{c}_j)$ for $j = 1, \cdots, p$, where Y_{ij}^* is a continuous variable underlying Y_{ij}. For the continuous outcomes, we have $Y_{ij} = Y_{ij}^*$, for $j = 1, \cdots, p_1$. For the discrete outcomes, with $Y_{ij} \in \{1, \cdots, d_j\}$, we have $Y_{ij} = \sum_{l=1}^{d_j} lI(c_{j,l-1} < Y_{ij}^* \leq c_{j,l})$ for $j = p_1+1, \cdots, p$, where $\mathbf{c}_j = (c_{j,0}, \cdots, c_{j,d_j})^T$ are the thresholds satisfying $-\infty = c_{j,0} < c_{j,1} < \cdots < c_{j,d_j} = \infty$, d_j is the number of categories of the jth outcome. Here, d_j can be close to ∞ as $n \to \infty$, therefore, 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 the following semiparametric linear transformation model of the form:

$$H_j(Y_{ij}^*) = \mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T e_i + \varepsilon_{ij}, \;\; j = 1, \ldots, p, \tag{2.1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \cdots, \boldsymbol{\beta}_p^T)^T$ is a vector of regression coefficients; $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)^T$ represent the factor loadings; e_i is a vector of latent variables summarizing the treatment effect for subject i; and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{ip})^T$ is a vector of independently error distributed as $N(0, \text{diag}(\sigma_1^2, \cdots, \sigma_p^2))$; $H_1, \cdots, H_p$ are the unknown increasing transformation functions, satisfying $H_j(-\infty) = -\infty$ and $H_j(\infty) = \infty$ for $j = 1, \cdots, p$. 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 denote $H_j(-\infty) = -\infty$ or $H_j(\infty) = \infty$. This is proper with the monotonicity of H_j.

Clearly, what distinguishes our model from the existing methods lies in the nonparametric link functions, $H_1, \ldots, H_p$, which 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 now relate the latent variable to $\mathbf{Z}_i$, which records treatment assignment and other covariates for the sake of comparisons, via

$$e_i = \gamma \mathbf{Z}_i + \epsilon_i, \tag{2.2}$$

where γ is an unknown regression coefficient matrix characterizing the treatment effect in a population, ϵ_i is the random error distributed as $N(0, \boldsymbol{\Sigma}_e)$, $\boldsymbol{\Sigma}_e = \text{diag}(\sigma_{e1}^2, \ldots, \sigma_{e,q}^2)$; here, $\mathbf{Z}_i$ and ϵ_i are 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 denote by $\tilde{\varepsilon}_{ij} = \alpha_j e_i + \varepsilon_{ij}$, and rewrite the model for the j-th

outcome as

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

Apparently, model (2.3) belongs to a rich family of semiparametric transformation models. For example, when H_j takes the form of a power function, model (2.3) reduces to the familiar Box-Cox transformation models (Box and Cox, 1964; Bickel and Doksum, 1981). If $H_j(y) = y$ and $H_j(y) = \log(y)$, model (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 et al. (2002), Horowitz (1996), Ye and Duan (1997), Chen (2002), Zhou, Lin and Johnson (2009) and Lin and Zhou (2009) 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 models (2.1) and (2.2). First, unobserved latent variables e_i are involved. Second, some outcomes, such as $Y_{ij}^*, j = p_1 + 1, \ldots, p$, are not completely observed. We address these issues in the next section.

3 Estimation

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

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

Hence, given ϵ_i, $H_1(Y_{i1}^*), \cdots, H_p(Y_{ip}^*)$ are independent and distributed as $H_j(Y_{ij}^*) \sim N(\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T \epsilon_i, \sigma_j^2)$ for $j = 1, \cdots, p$. For each given $j > p_1$, the discrete components, we can only estimate $H_j(c_{j,1}), \cdots, H_j(c_{j,d_j-1})$, as the $\mathbf{c}_j$ and H_j are

unidentifiable separately. To solve this problem, for each given $j > p_1$, we define a nondecreasing step function G_j with jumps only at $1, \cdots, d_j - 1$, and $G_j(m) = H_j(c_{j,m})$ for any $m \in \{1, \ldots, d_j - 1\}$, where $c_{j,m}$ is the unknown upper limit of Y_{ij}^* when $Y_{ij} = m$. To facilitate expression, we also denote $G_j = H_j$ for $j \leq p_1$, the part for the continuous outcome. The estimation of $H_j, j = 1, \ldots, p$ is thus transformed to the estimation of G_j for $j = 1, \cdots, p$.

Equations (3.1) continue to hold if H_j, $\boldsymbol{\beta}_j$, α_j and σ_j are replaced by H_j/c, $\boldsymbol{\beta}_j/c$, α_j/c and σ_j/c for any $c > 0$. Therefore, scale normalizations are needed to make identification possible. In the paper, we use $\sigma_j^2 = 1$, $j = 1, \cdots, p$ for scale indentification. In addition, we assume that $\mathbf{Z}_i$ and $\mathbf{X}_{ij}$ do not contain intercept term for location normalizations. As only $\boldsymbol{\alpha}\gamma$ and $\boldsymbol{\alpha}\boldsymbol{\Sigma}_e\boldsymbol{\alpha}^T$ are identifiable, further identification conditions are that $\sigma_{e,j}^2 = 1$ and $\alpha_{jk} = 0$ for all $j < k$, where $j = 1, \cdots, p$, $k = 1, \cdots, q$. Let $\boldsymbol{\Theta} = \{\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma\}$ and $\mathbf{G} = \{G_1, \cdots, G_p\}$, hence, $\boldsymbol{\Theta}$ and $\mathbf{G}$ are the unknown parameters and functions to be estimated in the semiparametric latent variable transformation models defined by (2.1) and (2.2).

3·1 *Estimations of the parameters* $\boldsymbol{\Theta}$

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

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

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

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

where $\Sigma_{22} = \boldsymbol{\alpha}\boldsymbol{\alpha}^T + I_{p\times p}$, the likelihood for the observed data can be expressed as

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

As the likelihood function involves the infinite dimensional parameter $G_j, j = 1, \cdots, p$, 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, $j = 1, \cdots, p$. Then, the parameter $\boldsymbol{\Theta}$ is estimated by maximizing a pseudo-likelihood, which is the likelihood function $L(\boldsymbol{\Theta}; \mathbf{G})$, with $\mathbf{G}$ being replaced by its estimated values. We repeat the procedure until convergence.

3·2 *Estimation of the transformation function*

We first estimate the transformation functions with a given $\boldsymbol{\Theta}$. For any given $j \leq p$, we consider $y_j \in \mathcal{R}$ if $j \leq p_1$ and $y_j \in \{1, \ldots, d_j\}$ for $j > p_1$, and the "marginal" probability for the event of $Y_{ij} \leq y_j$. It follows that

$$Pr(Y_{ij} \leq y_j | \mathbf{X}_{ij}, \mathbf{Z}_i) = Pr(H_j(Y_{ij}^*) \leq H_j(y_j) | \mathbf{X}_{ij}, \mathbf{Z}_i), \text{if j} \leq \text{p}_1$$

and

$$Pr(Y_{ij} \leq y_j | \mathbf{X}_{ij}, \mathbf{Z}_i) = Pr(H_j(Y_{ij}^*) \leq G_j(y_j) | \mathbf{X}_{ij}, \mathbf{Z}_i), \text{if j} > \text{p}_1,$$

both of which are equal to

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

under the convention of $G_j = H_j$ for $j \leq p_1$. Here $\phi(\cdot)$ denotes the density function for $q-$dimensional standard normal random vector and $\Phi(\cdot)$ the cumulative function

for the standard normal variable. This leads to a series of estimating equations

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

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

Due to the monotonicity of function Φ, it follows that the estimator $\widehat{G}_j(\cdot)$ of $G_j(\cdot)$ is a nondecreasing step function with jumps only at the observed $Y_{ij}, i=1,\cdots,n$, $j=1,\cdots,p$. Then solving the system of estimating equations of infinite number of equations defined by (3.3) is equivalent to solving the system of finite number of equations. In contrast with the traditional nonparametric approaches to estimating the transformation function (Horowitz, 1996; Zhou, Lin and Johnson, 2009), our approach does not involve nonparametric smoothing, and avoids smoothing related difficulties, in particular the selection of smoothing parameters.

Initial values are generally required for iteratively estimating $\boldsymbol{\Theta}$ and $G_j(\cdot)$, for which we propose the following procedure. Denote by $\gamma_j=\gamma^T\alpha_j$ for $j=1,\cdots,p$. A simple application of the double expectation theorem yields

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

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

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

equations

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

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

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

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

for $j=1,\ldots,p$. We set the starting values for $\alpha_j, j=1,\ldots,p$ 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 now present the large sample properties of the estimators derived in Section 3. Let $\widehat{\boldsymbol{\Theta}}$ and $\widehat{G}_j, j=1,\cdots,p$ denote the estimators of $\boldsymbol{\Theta}$ and $G_j, j=1,\cdots,p$. Throughout the article, we use the subscript "0" for the true value. For example, G_{j0} is the true value of G_j. Denote

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

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

$$d_j(y)=\frac{E\phi\left(\frac{G_{j0}(y)-W_{ij}(\boldsymbol{\Theta})}{\sqrt{\alpha_j^T\alpha_j+1}}\right)\left\{\frac{\partial W_{ij}(\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}+\left[G_{j0}(y)-W_{ij}(\boldsymbol{\Theta})\right]\frac{\partial\log(\sqrt{\alpha_j^T\alpha_j+1})}{\partial\boldsymbol{\Theta}}\right\}}{E\phi\left(\frac{G_{j0}(y)-W_{ij}(\boldsymbol{\Theta})}{\sqrt{\alpha_j^T\alpha_j+1}}\right)}|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}_0},$$

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

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

Theorem 1. When $n \to \infty$, $\widehat{\mathbf{\Theta}}$ and $\widehat{G}_j(y_j)$ are unique and uniformly consistent for $\mathbf{\Theta}_0$ and $G_{j0}(y_j)$ over $y_j \in [a_j, b_j]$ if $j \leq p_1$, and $y_j \in \{1, \ldots, d_j - 1\}$ if $j > p_1$.

Theorem 2. When $n \to \infty$, we have

$$n^{1/2}(\widehat{\mathbf{\Theta}} - \mathbf{\Theta}_0) \to N\left(0, \mathbf{B}^{-1}\mathbf{A}(\mathbf{B}^{-1})^T\right), \tag{4.1}$$

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

Theorem 3. When $n \to \infty$, we have

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

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

The results are interesting as $\widehat{G}_j(y)$ converges to $G_{j0}(y)$ at a rate of $n^{-1/2}$, implying that the nonparametric function $G_{j0}(\cdot)$ can be estimated with a parametric convergent rate. Similar conclusions but 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 $\widehat{\mathbf{\Theta}}$

As the involved computation prohibits the direct usage of the asymptotic variance of $\widehat{\mathbf{\Theta}}$ presented by Theorem 2, we propose to use a resampling scheme proposed by Jin et al. (2001) to evaluate the variability of $\widehat{\mathbf{\Theta}}$. Specifically, we first generate n exponential random variables $\xi_i, i = 1, \cdots, n$ 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 $\mathbf{\Theta}^*$ and $G_j^*(y), j = 1, \cdots, p$ for any y:

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

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

The estimates $\mathbf{\Theta}^*$ and $G_j^*(\cdot), j = 1, \cdots, p$ can be obtained using the same iterative algorithm described in Appendix A. Following Jin *et al.* (2001) and using the asymptotic expansion (8.12) in Appendix D, we establish the validity of the proposed resampling method.

Proposition Under the conditions given in Section 4, the conditional distribution of $n^{1/2}(\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}})$, given the observed data, converges almost surely to the asymptotic distribution of $n^{1/2}(\widehat{\mathbf{\Theta}} - \mathbf{\Theta}_0)$.

This result reveals that by repeatedly generating $\xi_1, \cdots, \xi_n$ many times, we can obtain a large number of realizations of $\mathbf{\Theta}^*$, the empirical variance of which can be used to approximate the variance of $\widehat{\mathbf{\Theta}}$.

6 Simulation

We examine the finite sample performance of the proposed method. Particularly, we investigate the robustness and the efficiency of the proposed method, in comparison with two "extreme" methods. The first method uses the models (2.1) and (2.2) with the misspecified transformation functions, and is acronymed the MT method. The

second method uses the models (2.1) and (2.2) with the correctly specified transformation functions and is termed the CT method. The joint normal models (JNM) essentially is the MT method. The MT estimator is used to investigate the robustness of the proposed method. The CT estimator is served as the gold standard that evaluates the efficiency of the proposed method. Finally, in each case we also evaluate the variance estimators described in Section 5. We assess the performance of the various estimators in terms of bias, standard deviation(SD) and the root of mean square error(RMSE).

Simulation 1 We simulated 500 datasets, each with 300 subjects. For each subject, the four outcomes $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$, where Y_{i1} and Y_{i2} are continuous, and Y_{i3} and Y_{i4} are discrete, are generated from the following transformation models

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

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

Table 1 presents 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 are based on 259 replications out of the 500 simulation runs, as the Newton-Raphson algorithm failed among 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. This occurs because the misspecification of H_1 and H_2 leads to the biased estimator of γ, which results in biased estimators of thresholds for the discrete responses (see Table 2), consequently, the parameters in the model for the discrete responses are biased. In contrast, our method yields estimates close to the true values, with variances that are very close to those for the CT estimators, suggesting that our procedure is robust with little loss of efficiency. We conjecture that this is largely due to the fact that the proposed estimation of the finite dimensional parameters is essentially MLE based. In addition, although the nonparameteric transformation function in the likelihood is estimated through estimate equations, it does not need smoothing and is still $\sqrt{n}$-consistent.

Table 1 is placed around here.

For each simulated dataset, we also obtain 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 is severely biased. In contrast, our proposed approach yields unbiased estimators with variances close to those of the CT estimators, reiterating that our method is robust and efficient. Figure 1 displays the averaged estimated transformation functions and their 95% empirical pointwise confidence limits based on the 500 simulated datasets, showing that the proposed estimates of the transformation functions are very close to the true transformation functions.

Table 2 and Figure 1 are placed around here.

We have also tested the accuracy of the estimation of the standard error given in §5. The standard deviations, denoted by SD in Tables 1 and 2, of the 500 estimated parameters, based on the 500 simulations, can be regarded as the true standard errors. To test the accuracy of the standard error estimator, we take three typical samples, which attained $25\%, 50\%$ and 75% of $ASE = \|\widehat{\mathbf{\Theta}} - \mathbf{\Theta_0}\|$, respectively, of the 500 simulations. The average of three estimated standard errors based on the 500 realizations of $\mathbf{\Theta}^*$, denoted by SE_{ave}, summarizes the overall performance of the standard error estimator. Table 3 shows that the performance of the standard error estimator is satisfactory.

Table 3 is placed around here.

Simulation 2 Our method requires the error term to be a Gaussian variable. To investigate the sensitivity of our method to such an assumption, we generate data according to the settings similar to those in the simulation 1 except that we take the two outcomes Y_{i1} and Y_{i3} and generate ε_{i1} and ε_{i3} from the centralized and scaled gamma distribution $(Gamma(\tau, 1) - \tau)/\sqrt{\tau}$, which approaches the standard normal when τ increases. We take $\tau = 100, 10, 5, 3$ and 1 to test the sensitivity of our method to the normal assumption. Table 4 presents the bias and SD for the parameters.

Table 4 is placed around here.

The results of the case with $\tau = 1$ marked by * are based on 418 replications as the algorithm failed to converge in 82 out of 500 simulations. A useful rule to evaluate the severity of bias, as suggested by Olsen & Schafer(2001), is to check whether the standardized bias (bias over standard deviation) exceeds 0.4. Accordingly, when $\tau \geq 10$, or both skewness and excess kurtosis are less than one, the proposed estimators

are nearly unbiased. When both skewness and excess kurtosis are around $1 \sim 2$, indicating that the error is away from Gaussian variable in moderate degree, the proposed estimators are acceptable although they are slightly biased. Only when both the skewness and excess kurtosis are larger than two and the error distribution becomes severely nonnormal, the estimators are biased.

7 Analysis of a Stroke Trial

We analyze a real 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 t-PA or placebo. Two primary outcome including the modified Rankin scale (RAN) and NIHSS were measured at three months after the trial began. RAN is a simplified overall assessment of function in which a score of 0 indicates the absence of symptoms and a score of 6, severe disability, while NIHSS, a measure of neurologic deficit, is on a continous 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), 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 among the two primary outcomes, namely, RAN and NIHSS, we fit the following the models,

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

where Y_1 is the NIHSS, a continuous outcome, and Y_2 is RAN, an ordinal outcome. Y_2^* is the underlying continuous variable for Y_2, and the link between the two variables is $Y_2 = \sum_{l=1}^{7}(l-1)I(c_{l-1} < Y_2^* < c_l)$, 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 1000 simulations. For comparison purposes, we also applied the traditional joint normal model (JNM), that is, the models (7.1) with H_1 and H_2 set to be linear functions, to the dataset. For the JNM method, about 50% of the 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 1000 replicates. Even with the biased repeated samples that favored the JNM method, our method yielded a much smaller p-values, suggesting that the proposed method maybe more parsimonious in detecting signals. To ascertain the proper transformation function, we displayed in Figure 2(a) the estimated transformation function and its 95% pointwise confidence limits.

Figure 2 is placed around here.

In addition, 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 the RAN; edema and Mass do not have significant effects on both NIHSS and RAN. The highly significant p-value (0.007) for γ showed that the disease condition is significantly improved after t-PA treatment. In contrast, the JNM method failed to detect the benefit of the t-PA

treatment with p=0.093. Indeed, our proposed method confirmed the results that the t-PA treatment is beneficial as published in the original report.

Tables 5 and 6 are placed around here.

Finally, we checked validity of the assumed semiparametric transformation model (7.1) by examining the agreement of the distribution of the estimated residual with that of the normal distribution. Figure 2(b) displays the plot of the empirical quantiles of the estimated residuals, defined by $\{\widehat{\varepsilon}_{i1} = \widehat{H}_1(Y_{i1}) - \mathbf{X}_i^T\widehat{\beta}_1, i = 1, \ldots, n\}$, 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 model (7.1).

Moreover, to see whether $\widehat{H}_1(y)$ is 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 $\{\tilde{\varepsilon}_{i2} = c\log(Y_{i1}) - X_i^T\widehat{\beta}_1, i = 1, \ldots, n\}$. 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 have developed a semiparametric latent variables normal transformation model to summarize the multiple correlated outcomes with generally continuous and discrete components. The theoretical studies show that our estimators are asymptotically normal with a convergent rate $n^{-1/2}$, which is comparable to the rate for a fully parametric regression model. The simulation studies show that the proposed method is robust with little loss of the efficiency. Analysis of a real world problem shows that the proposed method may shed some new insight on our understanding of a clinical problem.

We envision that we can extend our method 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, which have significantly improved this manuscript.

References

Bartholomew, D. J. and Knott, M. (1999) Latent Variable Models and Factor Analysis. *London: Arnold.*

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, *Journal of the American Statistical Association*, **76**, 296-311.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society*, B, **26**, 211-252.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British journal of mathematical & statistical psychology*,**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. *Journal of the American Statistical Association*, **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. *Scandinavian Journal of Statistics*, **15**, 1-23.

Doksum, K. A. (1987). An Extension of Partial Likelihood Methods for Proportional Hazard Models to General Transformation Models. *The Annals of Statistics*, **15**, 325-345.

Dunson, D.B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B.* , **62**, 355-366.

Dunson, D.B. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*, **98**, 555-563.

Fitzmaurice, G.M., and Laird, N.M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association*, **90**, 845-852.

Gueorguieva, R.V., and Agresti, A. (2001). A correlated probit model for joint modelling of clustered binary and continuous responses. *Journal of the American Statistical Association*, **96**, 1102-1112.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model, *Journal of Econometrics*, **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. *Journal of the Royal Statistical Society, Series 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. *Journal of the American Statistical Association*, **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 journal of mathematical and statistical psychology*, **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.

OBrien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics*, **40**, 1079-1087.

Olsen, M. K. & Schafer, J. (2001) A two-part random-effects model for semi-continuous longitudinal data. *Journal of the American Statistical Association*, **96**, 730-745.

Pocock, S. T., Geller, N. L., and Tsiatis, 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. *Statistics in 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. *Journal of the Royal Statistical Society, Series 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(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. *The Annals of Statistics*, **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. *Jour. Roy. Statist. Soc. 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.

Appendix A: Implementation

We outline the algorithm for estimating $\mathbf{\Theta}$ and $G_j(\cdot), j = 1, \cdots, p$ as follows:

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

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

$$\prod_{i=1}^{n} \int_x \prod_{j=1}^{p_1} \phi\left(G_j(Y_{ij}) - \left(\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T x\right)\right) \times \prod_{j=p_1+1}^{p} \left[\Phi\left(G_j(Y_{ij}) - \left(\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T x\right)\right) - \Phi\left(G_j(Y_{ij} - 1) - \left(\mathbf{X}_{ij}^T \boldsymbol{\beta}_j + \alpha_j^T \gamma \mathbf{Z}_i + \alpha_j^T x\right)\right)\right] \phi(x) dx, \quad (8.1)$$

which is a low dimension integration. Then, replacing the integral with the sampling mean, we estimate $\mathbf{\Theta}$ by maximizing the following likelihood,

$$\prod_{i=1}^{n}\sum_{k=1}^{R}\left\{\prod_{j=1}^{p_1}\phi\left(G_j(Y_{ij})-\left(\mathbf{X}_{ij}^T\boldsymbol{\beta}_j+\alpha_j^T\gamma\mathbf{Z}_i+\alpha_j^T y_k\right)\right)\right.$$
$$\times\prod_{j=p_1+1}^{p}\left[\Phi\left(G_j(Y_{ij})-\left(\mathbf{X}_{ij}^T\boldsymbol{\beta}_j+\alpha_j^T\gamma\mathbf{Z}_i+\alpha_j^T y_k\right)\right)\right.$$
$$\left.\left.-\Phi\left(G_j(Y_{ij}-1)-\left(\mathbf{X}_{ij}^T\boldsymbol{\beta}_j+\alpha_j^T\gamma\mathbf{Z}_i+\alpha_j^T y_k\right)\right)\right]\right\}, \quad (8.2)$$

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

- Step 2. Given $\mathbf{\Theta}$, we estimate $\mathbf{G}(\mathbf{y})$ at $\mathbf{y}=\mathbf{Y}_1,\cdots,\mathbf{Y}_n$ 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 $\mathbf{G}(\mathbf{y})$, denoted by $\widehat{\mathbf{G}}(\mathbf{y})$, are obtained by solving the equation (3.3) for $G_j(y_j), j=1,\cdots,p$ by replacing $\mathbf{\Theta}$ with its estimator from the iteration described here.

Appendix B: Notation

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

$$\psi(y_j)=E\phi\left(\frac{G_{j0}(y_j)-W_{ij}(\mathbf{\Theta}_0)}{\tilde{\sigma}_{j0}}\right),\ \ \xi_{ij}(y)=I\left(Y_{ij}\le y\right)-\Phi\left(\frac{G_{j0}(y)-W_{ij}(\mathbf{\Theta}_0)}{\tilde{\sigma}_{j0}}\right),$$

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

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

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

Two extra notations are needed to obtain the asymptotic normality for $\widehat{G}_j(y)$,

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

$$\mathbf{D}(y) = E\phi\left(\frac{G_{j0}(y) - W_{ij}(\mathbf{\Theta})}{\tilde{\sigma}_j}\right)\left\{[G_{j0}(y) - W_{ij}(\mathbf{\Theta})]\frac{\partial\tilde{\sigma}_j^{-1}}{\partial\mathbf{\Theta}} - \frac{\partial W_{ij}(\mathbf{\Theta})}{\tilde{\sigma}_j\partial\mathbf{\Theta}}\right\}|_{\mathbf{\Theta}=\mathbf{\Theta}_0}.$$

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 $y_j \in \mathcal{R} \equiv (-\infty, \infty), j = 1, \cdots, p$ and $\mathbf{\Theta} \in D_\eta = \{\mathbf{\Theta} : \|\mathbf{\Theta} - \mathbf{\Theta}_0\| \leq \eta\}$,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij} \leq y_j\right) - \Phi\left(\frac{G_{j0}(y_j) - W_{ij}(\mathbf{\Theta})}{\tilde{\sigma}_j} - \zeta\right)\right\}\\ &\quad\to E\left\{\Phi\left(\frac{G_{j0}(y_j) - W_{ij}(\mathbf{\Theta}_0)}{\tilde{\sigma}_{j0}}\right) - \Phi\left(\frac{G_{j0}(y_j) - W_{ij}(\mathbf{\Theta})}{\tilde{\sigma}_j} - \zeta\right)\right\}, \qquad (8.3)\end{aligned}$$

almost surely as $n \to \infty$, where $W_{ij}(\mathbf{\Theta}) = \mathbf{X}_{ij}^T\boldsymbol{\beta}_j + \alpha_j^T\gamma\mathbf{Z}_i$. The uniform convergence follows from the 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 indictor function class is VC and both the indictor 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 (8.3) follows from Van de Geer (2000).

Then it follows from (8.3) that for large n, $y_j \in \mathcal{R}$ and $\mathbf{\Theta} \in D_\eta$, and sufficiently large ζ,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij} \leq y_j\right) - \Phi\left(\frac{G_{j0}(y_j) - W_{ij}(\mathbf{\Theta})}{\tilde{\sigma}_j} - \zeta\right)\right\} > 0, \qquad (8.4)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij} \leq y_j\right) - \Phi\left(\frac{G_{j0}(y_j) - W_{ij}(\mathbf{\Theta})}{\tilde{\sigma}_j} + \zeta\right)\right\} < 0. \qquad (8.5)$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij} \leq y_j\right) - \Phi\left(\frac{\widehat{G}_j(y_j; \mathbf{\Theta}) - W_{ij}(\mathbf{\Theta})}{\tilde{\sigma}_j}\right)\right\} = 0. \qquad (8.6)$$

By differentiating both side of (8.6) with respect to $\boldsymbol{\Theta}$, we obtain the identity

$$\frac{\partial \widehat{G}_j(y_j;\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} = \frac{\sum_{i=1}^n \phi\left(\frac{\widehat{G}_j(y_j;\boldsymbol{\Theta})-W_{ij}(\boldsymbol{\Theta})}{\tilde{\sigma}_j}\right)\left\{\frac{\partial W_{ij}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} + \left[\widehat{G}_j(y_j;\boldsymbol{\Theta}) - W_{ij}(\boldsymbol{\Theta})\right]\frac{\partial \log \tilde{\sigma}_j}{\partial \boldsymbol{\Theta}}\right\}}{\sum_{i=1}^n \phi\left(\frac{\widehat{G}_j(y_j;\boldsymbol{\Theta})-W_{ij}(\boldsymbol{\Theta})}{\tilde{\sigma}_j}\right)}. \quad (8.7)$$

When $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0$, (8.4) and (8.5) hold for any $\zeta > 0$, we have that $\widehat{G}_j(y_j;\boldsymbol{\Theta}_0) \to G_0(y_j)$ uniformly in $y_j \in \mathcal{R}$. Hence

$$\frac{\partial \widehat{G}_j(y_j;\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}} \to \frac{E\phi\left(\frac{G_{j0}(y_j)-W_{ij}(\boldsymbol{\Theta})}{\tilde{\sigma}_j}\right)\left\{\frac{\partial W_{ij}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} + [G_{j0}(y_j) - W_{ij}(\boldsymbol{\Theta})]\frac{\partial \log \tilde{\sigma}_j}{\partial \boldsymbol{\Theta}}\right\}}{E\phi\left(\frac{G_{j0}(y_j)-W_{ij}(\boldsymbol{\Theta})}{\tilde{\sigma}_j}\right)}|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}_0} \widehat{=} d_j(y_j). \quad (8.8)$$

To show the existence and uniqueness of $\widehat{\boldsymbol{\Theta}}$, we let $W(\boldsymbol{\Theta};\mathbf{G}) = \frac{\partial \log L(\boldsymbol{\Theta};\mathbf{G})}{\partial \boldsymbol{\Theta}}$, and $S(\boldsymbol{\Theta}) = \frac{1}{n}W(\boldsymbol{\Theta};\widehat{\mathbf{G}}(\boldsymbol{\Theta}))$, which is $W(\boldsymbol{\Theta};\mathbf{G})$ with $G_j(\cdot), j = 1,\cdots,p$ replaced by $\widehat{G}_j(\cdot;\boldsymbol{\Theta}), j = 1,\cdots,p$. It follows from (8.7) and $\widehat{G}_j(y_j;\boldsymbol{\Theta}_0) \to G_{j0}(y_j)$ uniformly in $y_j \in \mathcal{R}$ that

$$\frac{\partial S(\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}^T} = \frac{1}{n}\left\{\frac{\partial W(\boldsymbol{\Theta};\mathbf{G})}{\partial \boldsymbol{\Theta}^T} + \sum_{i=1}^n\left(\sum_{j=1}^p \frac{\partial^2 \log L_i(\boldsymbol{\Theta}_0;\mathbf{G})}{\partial \boldsymbol{\Theta}\partial G_j(Y_{ij})}\frac{\partial \widehat{G}_j(Y_{ij};\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}^T} + \sum_{j=p_1+1}^p \frac{\partial^2 \log L_i(\boldsymbol{\Theta}_0;\mathbf{G})}{\partial \boldsymbol{\Theta}\partial G_j(Y_{ij}-1)}\frac{\partial \widehat{G}_j(Y_{ij}-1;\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}^T}\right)\right\}|_{\mathbf{G}=\widehat{\mathbf{G}}(\boldsymbol{\Theta}),\boldsymbol{\Theta}=\boldsymbol{\Theta}_0} \to \mathbf{B},$$

where $\mathbf{B}$ is defined in Section 4. Now, because $S(\boldsymbol{\Theta}_0) \to 0$ and $\mathbf{B}$ is negative definite, there exists a unique solution $\widehat{\boldsymbol{\Theta}}$ to the equation $S(\boldsymbol{\Theta}) = 0$ in a neighborhood of $\boldsymbol{\Theta}_0$. The foregoing proof also implies that $\widehat{\boldsymbol{\Theta}}$ is strong consistent and that $\widehat{G}_j(y_j) = \widehat{G}_j(y_j;\widehat{\boldsymbol{\Theta}}) \to G_{j0}(y_j)$ almost surely uniformly in $y_j \in \mathcal{R}$. Thus Theorem 1 is completed.

Appendix D: Proof of Theorem 2

By the consistency of $\widehat{\boldsymbol{\Theta}}$ and a Taylor series expansion of $S(\widehat{\boldsymbol{\Theta}})$ around $\boldsymbol{\Theta}_0$, we get

$$\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0 \approx -\mathbf{B}^{-1} S(\boldsymbol{\Theta}_0). \tag{8.9}$$

Note that

$$\begin{aligned} S(\boldsymbol{\Theta}_0) &= \left\{ n^{-1} \frac{\partial \log L(\boldsymbol{\Theta}_0; \mathbf{G}_0)}{\partial \boldsymbol{\Theta}} + n^{-1} \frac{\partial \log L(\boldsymbol{\Theta}_0; \widehat{\mathbf{G}}(\boldsymbol{\Theta}_0))}{\partial \boldsymbol{\Theta}} - n^{-1} \frac{\partial \log L(\boldsymbol{\Theta}_0; \mathbf{G}_0)}{\partial \boldsymbol{\Theta}} \right\} \\ &\approx n^{-1} \frac{\partial \log L(\boldsymbol{\Theta}_0; \mathbf{G}_0)}{\partial \boldsymbol{\Theta}} + n^{-1} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{p} \frac{\partial \log L_i(\boldsymbol{\Theta}_0; \mathbf{G}_0)}{\partial \boldsymbol{\Theta} \partial G_j(Y_{ij})} \left(\widehat{G}_j(Y_{ij}; \boldsymbol{\Theta}_0) - G_{j0}(Y_{ij}) \right) \right. \\ &\qquad \left. + \sum_{j=p_1+1}^{p} \frac{\partial \log L_i(\boldsymbol{\Theta}_0; \mathbf{G}_0)}{\partial \boldsymbol{\Theta} \partial G_j(Y_{ij}-1)} \left(\widehat{G}_j(Y_{ij}-1; \boldsymbol{\Theta}_0) - G_{j0}(Y_{ij}-1) \right) \right\}. \end{aligned} \tag{8.10}$$

Because (8.6), we have

$$\begin{aligned} \widehat{G}_j(y_j; \boldsymbol{\Theta}_0) - G_{j0}(y_j) = \frac{\tilde{\sigma}_{j0}}{n\psi(y_j)} \sum_{i=1}^{n} \left\{ I\left(Y_{ij} \le y_j\right) - \Phi\left(\frac{G_{j0}(y_j) - W_{ij}(\boldsymbol{\Theta}_0)}{\tilde{\sigma}_{j0}} \right) \right\} \\ + o_p(n^{-1/2}), \end{aligned} \tag{8.11}$$

where $\psi(y_j)$ is defined in Section 4. Substituting (8.11) into (8.10) and exchanging the summations, we get

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

Hence, by (8.9), w have

$$\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0 \approx -n^{-1} \mathbf{B}^{-1} \sum_{i=1}^{n} \left\{ \frac{\partial \log L_i(\boldsymbol{\Theta}_0; \mathbf{G}_0)}{\partial \boldsymbol{\Theta}} + \sum_{j=1}^{p} \varphi_{ij1} + \sum_{j=p_1+1}^{p} \varphi_{ij2} \right\}. \tag{8.12}$$

The proof of Theorem 2 is completed.

Appendix E: Proof of Theorem 3

Because (8.6), for any $y \in \mathcal{R}$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij}\le y\right)-\Phi\left(\frac{G_{j0}(y)-W_{ij}(\boldsymbol{\Theta}_0)}{\tilde{\sigma}_{j0}}\right)\right\}\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\left\{\Phi\left(\frac{G_{j0}(y)-W_{ij}(\boldsymbol{\Theta}_0)}{\tilde{\sigma}_{j0}}\right)-\Phi\left(\frac{G_{j0}(y)-W_{ij}(\widehat{\boldsymbol{\Theta}})}{\sqrt{\widehat{\alpha}_j^T\alpha_j+1}}\right)\right\}\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\left\{\Phi\left(\frac{G_{j0}(y)-W_{ij}(\widehat{\boldsymbol{\Theta}})}{\sqrt{\widehat{\alpha}_j^T\alpha_j+1}}\right)-\Phi\left(\frac{\widehat{G}_j(y)-W_{ij}(\widehat{\boldsymbol{\Theta}})}{\sqrt{\widehat{\alpha}_j^T\alpha_j+1}}\right)\right\}=0,
\end{aligned}$$

hence

$$\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}(y)-\mathbf{D}^T(y)\left(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0\right)-\frac{\psi(y)}{\tilde{\sigma}_{j0}}\left(\widehat{G}_j(y)-G_{j0}(y)\right)=o_p(n^{-1/2}),$$

where $\mathbf{D}(y)$ is defined in Appendix B. Substituting (8.12) into the equation above, we obtain,

$$\begin{aligned}
\widehat{G}_j(y)-G_{j0}(y)=\frac{\tilde{\sigma}_{j0}}{n\psi(y)}\sum_{i=1}^{n}\Big\{&\xi_{ij}(y)+\mathbf{D}^T(y)\mathbf{B}^{-1}\\
&\times\left(\frac{\partial\log L_i(\boldsymbol{\Theta}_0;\mathbf{G}_0)}{\partial\boldsymbol{\Theta}}+\sum_{j=1}^{p}\varphi_{ij1}+\sum_{j=p_1+1}^{p}\varphi_{ij2}\right)\Big\}+o_p(n^{-1/2}).
\end{aligned}\tag{8.13}$$

The proof of Theorem 3 is completed.

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

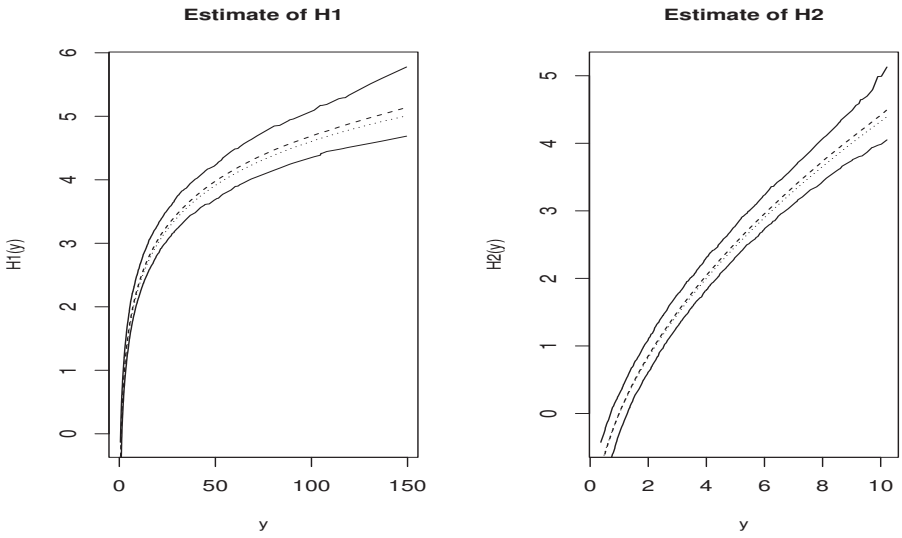

Figure 1: The estimated transformation functions (dotted-lined—true function; solid—95% confidential limit; dashed—average of the estimated transformation function).

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			

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

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 1000 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 1000 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

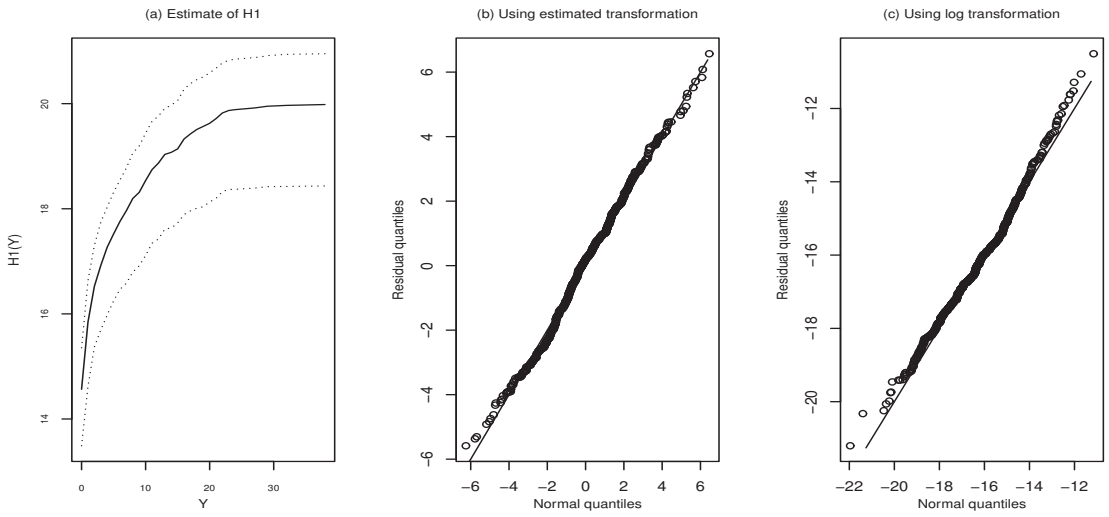

Figure 2: (a) The estimate (Solid) and its 95% confidence limits (dashed) of the transformation function H_1 for the NIHSS; (b) The empirical quantiles of the estimated residuals $\{\widehat{\varepsilon}_{i1}\}$ against the normal theoretical quantiles when the transformation functions are estimated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