Analysis of Correlated Binary Data under Partially Linear Logistic Models

Grace Y. Yi

Department of Statistics and Actuarial Science, University of Waterloo, Canada N2L 3G1

Wenqing He

Department of Statistical and Actuarial Sciences, University of Western Ontario, Canada N6A 5B7

Hua Liang

Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY 14642, USA

Abstract

Clustered data arise commonly in practice and it is often of interest to estimate the mean response parameters as well as the association parameters. However, most research has been directed to address the mean response parameters with the association parameters relegated to a nuisance role. There is little work concerning both the marginal and association structures, especially in the semiparametric framework. In this paper, our interest centers on inference on both the marginal and association parameters, which is driven by many practical applications. We develop semi-parametric methods for clustered binary data and establish the theoretical results. The proposed methodology is investigated through various numerical studies.

Key Words and Phrases: Association; binary outcomes; clustered data; estimating equation; semiparametric estimation.
1 Introduction

Clustering data, including longitudinal and multivariate data, arise frequently in health and medical studies. These data may occur when subsampling the primary sampling units or repeatedly collecting measurements over time for subjects in the study. As is well-known, standard univariate analysis methods may not be suitable to handle clustered data because individuals in the same cluster cannot be treated as functionally independent. The challenge arising from analyzing such data centers on dealing with association among units or subjects within a cluster.

Typically, marginal methods such as generalized estimating equations (GEE) techniques are commonly utilized to analyze clustered data. The methods emphasize modulating the mean structure with the full distribution of the responses left unspecified, and thus the methods are generally viewed attractive. It is often the case to assume that the mean response conforms to a generalized linear model, where covariates pertaining to the response are presented in a linear form through a link function. Under this setup properly facilitating association structures may increase the efficiency of estimation of the marginal response parameters (e.g., Prentice 1988; Yi and Cook 2002); whereas ignoring association between responses or incorrectly assumed correlation structure may lead to biased variance estimates and thus biased inference (e.g., Zeger and Liang 1986).

In practice, however, the relationship between the response and covariates may be very complex and linear terms may not be adequate enough to feature that relationship. It may be even worse than useless to fit a linear model to a nonlinear relationship sometimes. Under these circumstances, a semiparametric regression with both a linear term $x'\beta$ and a nonlinear term $\theta(z'\alpha)$ included may be preferable, where $\theta(.)$ is a smooth but unknown function. Furthermore, it is a common practice to induce a nonparametric function into the model for covariates that have a large dimension and are of little interest (e.g., confounders). This enables us to make inference on the effects of $x$ while making minimal assumptions on $z$ with a nonparametric function $\theta(.)$.

Marginal semiparametric models based on using GEE methods and their various extensions have become increasingly popular. See, for example, Severini and Staniswalis (1994), Carroll, Fan, Gijbels, and Wand (1997), Xia, Tong, and Li (1999), Lin and Carroll (2001a,b), Wang (2003), Fan and Li (2004), Wang, Carroll and Lin (2005), and Xia and Härdle (2006), among others. However, they concerned the marginal mean parameters only with the association parameters treated as
nuisance.

In many practical applications simply working on the marginal mean responses could be very restrictive. Estimation of the association parameters may be the central theme of the study. For example, in familial studies of inherited traits (e.g., Galton 1889; McCullagh and Nelder 1989) and developmental toxicology studies of laboratory animals (e.g., Hall and Severini 1998), subjects in a family or cluster share common genetic traits or are subject to common environmental factors, and hence it is of prime scientific interest to study the association between responses.

In the literature, however, there is little discussion on featuring both the mean and association structures with semiparametric regression modeling. To fill up this gap, in this paper, our objective is to elucidate association structures for clustered data, in conjunction with modeling the marginal mean responses. We develop inference procedures for clustered data. The discussion focuses on binary responses, which is driven by the fact that binary outcomes often arise in distinct contexts. We specifically investigate semiparametric regressions which make it possible to study a richer class of mean and association structures with more complex relationships. However, such a flexibility presents considerable challenges in estimation procedures. The computing algorithm for usual estimating equations based on the Newton-Raphson method can not be employed directly due to the inclusion of a nonlinear function whose form is not known. To circumvent this problem, we use the local polynomial smoothing technique to perform estimation, and the discussion may be readily adapted to other approach such as smoothing spline estimation. Moreover, the inclusion of a nonparametric term $\theta(\cdot)$ into the mean model in combination with modeling the association structure makes it very difficult to establish the asymptotic results.

The remainder of the article is organized as follows. The notation and inference framework are introduced in Section 2 and estimation procedures are described in Section 3. In Section 4 we establish the asymptotic properties of the resulting estimators and discuss the issue of parameter interpretation. Numerical studies are given in Section 5 to illustrate the proposed methods and to assess its performance under a variety of settings. We conclude the article with a discussion in the last section.
2 Notation and Framework

Suppose that there are \( n \) clusters and \( m_i \) subjects within cluster \( i \), \( i = 1, \ldots, n \). Let \( Y_{ij} \) be the binary response for subject \( j \) in cluster \( i \), \( x_{ij} \) and \( z_{ij} \) be the \( p \times 1 \) and \( q \times 1 \) covariate vectors, respectively. Denote \( Y_i = (Y_{i1}, \ldots, Y_{im_i})^T \). Denote \( x_i = (x_{i1}, \ldots, x_{im_i})^T \) and \( z_i = (z_{i1}, \ldots, z_{im_i})^T \).

Define \( \mu_{ij} = E(Y_{ij} | x_i, z_i) \), and let \( \mu_i = (\mu_{i1}, \ldots, \mu_{im_i})^T, i = 1, \ldots, n \). Provided the mean of \( Y_{ij} \) depends only on the covariate vector for subject \( j \), i.e. \( E(Y_{ij} | x_i, z_i) = E(Y_{ij} | x_{ij}, z_{ij}) \) (Pepe and Anderson 1994), we consider the regression model

\[
\logit \mu_{ij} = x_{ij}^T \beta + \theta(z_{ij}^T \alpha) \quad \text{with } \|\alpha\| = 1
\]

(2.1)

where \( \beta \) and \( \alpha \) are \( p \times 1 \) and \( q \times 1 \) unknown parameter vectors, respectively, and \( \theta(\cdot) \) is an unknown smoothing function. Other monotone link functions, such as probit, or complementary log-log functions, may also be used to replace the logit link if necessary. The requirement of \( \|\alpha\| = 1 \) ensures identifiability of \( \alpha \).

We assume that \( Y_{ij} \) and \( Y_{ij'} \) are independent for different clusters \( i \) and \( i' \). But within the same cluster, the responses may be correlated. Various measures have been proposed to quantify the association between binary outcomes. For example, Prentice (1988) discussed using correlation coefficients for measuring association for longitudinal binary data, and Zhao and Prentice (1990) discussed a measure based on covariances. Odds ratios, on the other hand, have received increasing interest due to the fact that there is no constraint associated with such measures. Specifically, conditional odds ratios (e.g., Fitzmaurice and Laird 1993) and marginal odds ratios (e.g., Lipsitz, Laird, and Harrington 1991) have been widely used. As conditional odds ratios may not have a convenient interpretation that is independent of the cluster size, in this paper we focus discussion on marginal odds ratios. Let \( \psi_{ijk} \) be the odds ratio between responses \( Y_{ij} \) and \( Y_{ik} \) in cluster \( i \) \( (j < k) \), defined by

\[
\psi_{ijk} = \frac{P(Y_{ij} = 1, Y_{ik} = 1 | x_i, z_i) \cdot P(Y_{ij} = 0, Y_{ik} = 0 | x_i, z_i)}{P(Y_{ij} = 1, Y_{ik} = 0 | x_i, z_i) \cdot P(Y_{ij} = 0, Y_{ik} = 1 | x_i, z_i)}
\]

Regression models may be employed to feature various association structures, with the dependence of the association on covariates being explicitly reflected. Typically, a log-linear regression may be assumed with \( \log \psi_{ijk} = u_{ijk}^T \Phi \), where \( u_{ijk} \) is a vector of covariates which specifies the
form of the association between \( Y_{ij} \) and \( Y_{ik} \), and \( \phi \) is a vector of regression parameters. Letting \( u_{i,j,k} \) be the scalar one, for instance, leads to the exchangeable association between responses within the same cluster; setting \( u_{i,j,k}^T \phi = \phi^{[j-k]} \) results in an autoregressive correlation among responses \((j < k)\).

Let \( \mu_{i,j,k} = P(Y_{ij} = 1, Y_{ik} = 1|x_i, z_i) \) be the joint probability for the pair \( (Y_{ij}, Y_{ik}) \), given the covariates \( x_i \) and \( z_i \). It is determined by the marginal means and the odds ratio, given by (e.g., Lipsitz et al.1991; Yi and Thompson 2005)

\[
\mu_{i,j,k} = \begin{cases} 
\frac{a_{i,j,k} - \{a_{i,j,k}^2 - 4\psi_{i,j,k}(\psi_{i,j,k} - 1)\mu_{ij}\mu_{ik}\}^{1/2}}{2(\psi_{i,j,k} - 1)}, & \text{if } \psi_{i,j,k} \neq 1 \\
\mu_{ij}\mu_{ik}, & \text{if } \psi_{i,j,k} = 1
\end{cases}
\]

where \( a_{i,j,k} = 1 - (1 - \psi_{i,j,k})(\mu_{ij} + \mu_{ik}) \).

## 3 Estimation Procedures

In this section we describe marginal methods for estimating mean response parameters \( \alpha \) and \( \beta \) and association parameters \( \phi \). Let \( V_i = [v_{i,j}] \) be the true covariance matrix for the response vector \( Y_i \) for cluster \( i \), with \( v_{i,j,j} = \mu_{ij}(1 - \mu_{ij}) \) and \( v_{i,j,k} = \mu_{i,j,k} - \mu_{ij}\mu_{ik} \) for \( j \neq k \), and \( W_i = \text{diag}(\sqrt{v_{i,j,j}}, j = 1, ..., m_i)C_i\text{diag}(\sqrt{v_{i,j,j}}, j = 1, ..., m_i) \) be a working matrix, where \( C_i \) is an invertible working correlation matrix. Throughout the paper we assume that \( C_i \) may depend on a parameter vector that is distinct from the mean response parameters \( \alpha \) and \( \beta \).

Let \( U_{i,\alpha}(\alpha, \beta, \theta(\cdot)) = \left(\frac{\partial \mu^T}{\partial \alpha}\right) W_i^{-1}(y_i - \mu_i) \) and \( U_{i,\beta}(\alpha, \beta, \theta(\cdot)) = \left(\frac{\partial \mu^T}{\partial \beta}\right) W_i^{-1}(y_i - \mu_i) \). It can be seen that both \( U_{i,\alpha}(\alpha, \beta, \theta(\cdot)) \) and \( U_{i,\beta}(\alpha, \beta, \theta(\cdot)) \) have zero expectation, i.e., they are unbiased estimating functions for \( \alpha \) and \( \beta \).

To estimate the association parameters \( \phi \), one may conduct estimation in the same spirit of Prentice (1988) by constructing a second set of unbiased estimating functions based on the method of moments. As the cluster size increases, this approach may become computationally burdensome. Alternatively, we may employ the alternating logistic regression discussed in Carey, Zeger, and Diggle (1993) where the conditional expectation \( \xi_{i,j,k} = E(Y_{ij}|Y_{ik} = y_{ik}, x_i, z_i) \) is needed for
$j < k$. The conditional expectation $\xi_{i;jk}$ is related to the association, marginal and joint probabilities by the following expression:

$$\xi_{i;jk} = \expit \left\{ (\log \psi_{i;jk})y_{ik} + \log \left( \frac{\mu_{ik} - \mu_{i;jk}}{1 - \mu_{ij} - \mu_{ik} + \mu_{i;jk}} \right) \right\}$$

where $\expit(t) = \exp(t)/(1 + \exp(t))$. Let $W_i^* = \text{diag}\{\xi_{i;jk}(1 - \xi_{i;jk}), j < k\}$ be the working matrix, then $U_i^\varphi(\alpha, \beta, \theta(\cdot), \varphi) = \left( \frac{\partial \xi_{i;jk}^\varphi}{\partial \varphi} \right) W_i^{*-1} \epsilon_i$, are unbiased estimating functions for $\varphi$, where $\epsilon_i = (y_{i1} - \xi_{i;12}, \ldots, y_{i1} - \xi_{i;1m}, y_{i2} - \xi_{i;23}, \ldots, y_{i,m-1} - \xi_{i;m-1,m})^T$, and $\xi_i = (\xi_{i;12}, \ldots, \xi_{i;1m}, \xi_{i;23}, \ldots, \xi_{i;m-1,m})^T$.

If $\theta(\cdot)$ is known to be a linear function, then estimation of $\alpha$, $\beta$ and $\phi$ may proceed in a straightforward manner, as outlined in Carey et al. (1993) and Yi and Cook (2002), where the working matrix $W_i$ may be taken as the true covariance matrix $V_i$. Since the estimating functions $U_i^\alpha(\alpha, \beta, \theta(\cdot)), U_i^\beta(\alpha, \beta, \theta(\cdot))$ and $U_i^\varphi(\alpha, \beta, \theta(\cdot), \phi)$ involve an unknown smooth function $\theta(\cdot)$, we need to use nonparametric approaches to estimate this function locally in order to estimate $\alpha$, $\beta$, and $\phi$. Assuming $\theta(u)$ has the second derivative, we may approximate $\theta(u)$ by a locally linear function within the neighborhood of $u_0$ via the Taylor series expansion $\theta(u) \approx \theta(u_0) + \theta'(u_0)(u - u_0)$ for a given point $u_0$. Denote $a_0(u_0) = \theta(u_0), a_1(u_0) = \theta'(u_0)$, and $a(u_0) = (a_0(u_0), a_1(u_0))^T$.

We further introduce the following notation: $U_{ij} = z_{ij}^\alpha, \gamma_{ij}(u; \alpha) = (1, U_{ij} - u), \Gamma_i(u)$ is an $m_i \times 2$ matrix with the $j$th row being $\gamma_{ij}(u; \alpha); \Delta_i = \text{diag}(\mu_{ij}^{(1)}(\cdot)) = 1, \ldots, m_i$, where $\mu_{ij}^{(1)}(\cdot)$ is the first derivative of the function $\mu_{ij}(\cdot)$ evaluated at $x_{ij}^\alpha \beta + \theta(z_{ij}^\alpha \alpha)$). Let $K(u)$ be a kernel function (or a symmetric density function) with a compact support and $h$ be a bandwidth. Denote $K_h(t) = K(t/h)/h$. Let $K_{ih}(u) = \text{diag}(K_{ih}(U_{ij} - u), j = 1 \ldots, m_i)$ be a diagonal matrix. Below we describe a two-stage algorithm for estimation of mean parameters $\alpha$ and $\beta$ and association parameter $\phi$.

In stage 1 we adopt the working independence matrix to conduct estimation of $\alpha$ and $\beta$ using the local linear profile kernel method. In stage 2 we estimate association parameter $\phi$ using the usual GEE approach. This estimation strategy stems from the fact that use of the working independence matrix allows us to ignore the association parameters temporarily yet to lead to a consistent estimator for the marginal mean parameters $\alpha$ and $\beta$ (Liang and Zeger 1986; Lin and Carroll 2001a).

Stage 1:
Step 0. Choose initial values $\alpha_0$ and $\beta_0$, and set $\hat{\alpha} = \alpha_0 / \|\alpha_0\|$ and $\hat{\beta} = \beta_0$.

Step 1. For a given point $u$ in a selected grid find $\hat{\theta}(u; \hat{\alpha}, \hat{\beta}) = \tilde{a}_0(u)$ by solving the following equations

$$
\sum_{i=1}^{n} \Gamma_i^T(u) \Delta_i \tilde{K}_{ih}(u) \tilde{W}_i^{-1}(y_i - \bar{\mu}_i) = 0 \tag{3.2}
$$

with respect to $a(u)$, where $\tilde{K}_{ih}(u)$ is $K_{ih}(u)$ with $\alpha$ replaced by $\hat{\alpha}$, $\bar{\mu}_i = (\bar{\mu}_{i1}, \ldots, \bar{\mu}_{im_i})^T$, $\bar{\mu}_{ij} = g(x_{ij}^T \beta + \gamma_{ij}(u; \alpha) a(u)) = g(x_{ij}^T \beta + a_0(u) + a_1(u)(U_{ij} - u))$ with $\beta$ and $\alpha$ replaced by $\hat{\beta}$ and $\hat{\alpha}$, respectively, $\Delta_i$ is $\Delta_i$ with $\mu_{ij}$ replaced by $\bar{\mu}_{ij}$, $g(t) = \expit(t)$, and $\tilde{W}_i$ is the working independence matrix $\text{diag}\{\bar{\mu}_{ij}(1 - \bar{\mu}_{ij}), j = 1, ..., m_i\}$.

Step 2. Given the estimate $\hat{\theta}(u; \hat{\alpha}, \hat{\beta}) = \tilde{a}_0(u)$ and $\tilde{a}_1(u)$ for points $u$ in the selected grid, update $(\hat{\alpha}, \hat{\beta})$ by solving the following equations for $\alpha$ and $\beta$:

$$
\sum_{i=1}^{n} \frac{\partial \tilde{\mu}_i^T(\alpha, \beta)}{\partial \alpha} \tilde{W}_i^{-1}\{y_i - \bar{\mu}_i(\alpha, \beta)\} = 0 \tag{3.3}
$$

$$
\sum_{i=1}^{n} \frac{\partial \tilde{\mu}_i^T(\alpha, \beta)}{\partial \beta} \tilde{W}_i^{-1}\{y_i - \bar{\mu}_i(\alpha, \beta)\} = 0 \tag{3.4}
$$

where $\tilde{\mu}_i(\alpha, \beta) = (\bar{\mu}_{i1}(\alpha, \beta), \ldots, \bar{\mu}_{im_i}(\alpha, \beta))^T$ with $\bar{\mu}_{ij}(\alpha, \beta) = g(x_{ij}^T \beta + \hat{\theta}(z_{ij}^T \alpha; \hat{\alpha}, \hat{\beta}))$, and $\tilde{W}_i = \text{diag}(w_{ij}, j = 1, ..., m_i)$ with $w_{ij} = g(x_{ij}^T \beta + \tilde{a}_0(z_{ij}^T \alpha) + \tilde{a}_1(z_{ij}^T \alpha)(z_{ij}^T \alpha - z_{ij}^T \hat{\alpha}))$.

Step 3. Repeat steps 1 and 2 until convergence of $(\hat{\alpha}, \hat{\beta})$.

Stage 2: To estimate the association parameter $\phi$, we solve the following equations:

$$
\sum_{i=1}^{n} U_{i\phi}(\hat{\alpha}, \hat{\beta}, \hat{\theta}(z_{ij}^T \alpha), \phi) = 0 \tag{3.5}
$$

with respect to $\phi$, where $U_{i\phi}(\hat{\alpha}, \hat{\beta}, \hat{\theta}(z_{ij}^T \alpha), \phi)$ is $U_{i\phi}(\alpha, \beta, \theta(\cdot), \phi)$ with $\mu_{ij}$ replaced by $g(x_{ij}^T \beta + \hat{\theta}(z_{ij}^T \alpha))$, and $\theta(\cdot)$ replaced by $\hat{\theta}(z_{ij}^T \alpha) = \tilde{a}_0(z_{ij}^T \alpha)$ which is the value of $\tilde{a}_0(\cdot)$ obtained at the convergence of $(\hat{\alpha}, \hat{\beta})$. Denote by $\hat{\phi}$ the resulting estimate of $\phi$. 

6
4 Theory

4.1 Asymptotic Properties

Here we establish the asymptotic properties for the resultant estimators \((\hat{\alpha}, \hat{\beta})\) and \(\hat{\phi}\). Analogous to those of Lin and Carroll (2001a,b) and Wang (2003) we work with the case with \(m_i \equiv m\) for ease of notation. Covariates \(x_i\) and \(z_i\) are allowed to be correlated. The triples \((Y_i, x_i, z_i), i = 1, 2, \ldots, n\), are assumed independently identically distributed. Both \(V_i\) and \(C_i\) are assumed invertible.

Let \(Q_{11}(u) = \sum_{s=1}^{m} E\{[\tilde{\mu}_{1s}^{(1)}]^2 w_i^{s} | U_{1s} = u\}\), where \(w_i^{s} = 1/(\tilde{\mu}_{is}(1 - \tilde{\mu}_{is}))\) is the \(s\)th diagonal element of the inverse matrix \(\tilde{W}_1^{-1}\). Let \(\theta'_{\text{diag}}(z_i, \alpha) = \text{diag}(\theta'(z_i^T \alpha), j = 1, \ldots, m)\) and \(Q_{\text{diag}}(z_i, \alpha) = \text{diag}(Q_{11}(z_{ij}^T \alpha), j = 1, \ldots, m)\) be \(m \times m\) diagonal matrices. \(Q_x(u) = \sum_{s=1}^{m} E\{[\tilde{\mu}_{1s}^{(1)}]^2 w_i^{s} x_i^T | U_{1s} = u\}\), and \(Q_z(u) = \sum_{s=1}^{m} E\{[\tilde{\mu}_{1s}^{(1)}]^2 w_i^{s} \theta'(u) z_i^T | U_{1s} = u\}\). Define \(Q_x(z_i, \alpha) = \{Q_x^T(z_{i1}^T \alpha), \ldots, Q_x^T(z_{im}^T \alpha)\}^T\), and \(Q_z(z_i, \alpha) = \{Q_z^T(z_{i1}^T \alpha), \ldots, Q_z^T(z_{im}^T \alpha)\}^T\). Let \(\Lambda_i = \left(\begin{array}{c} z_i^T \theta'_{\text{diag}}(z_i, \alpha) \\ x_i^T \end{array}\right)\), \(A_i = \Lambda_i \Delta_i W_i^{-1} \Delta_i Q_{\text{diag}}^{-1}(z_i, \alpha)\), \(P_i = \Lambda_i + \{E(A_i 1_m | U_{i1}), \ldots, E(A_i 1_m | U_{im})\}\), and

\[
B_i = \Lambda_i \Delta_i W_i^{-1} \Delta_i \begin{pmatrix} \theta'_{\text{diag}}(z_i, \alpha) z_i + Q_{\text{diag}}^{-1}(z_i, \alpha) Q_z(z_i, \alpha) \\ x_i + Q_{\text{diag}}^{-1}(z_i, \alpha) Q_x(z_i, \alpha) \end{pmatrix}^T,
\]

where \(1_r\) is a \(r \times 1\) unit vector. If \(B = E(B_i | U_i)\), and \(\Sigma = E(P_i \Delta_i W_i^{-1} V_i W_i^{-1} \Delta_i^T \Sigma P_i^T)\), where \(U_i = (U_{i1}, \ldots, U_{im})^T\), then we have the following asymptotic result.

**Theorem 4.1** Under the conditions in Appendix 1, as \(nh^4 \to 0\) and \(nh^2 / \log(1/h) \to \infty\),

\[
\sqrt{n} \left\{ (\hat{\alpha} - \alpha)^T, (\hat{\beta} - \beta)^T \right\}^T
\]

is asymptotically normal with mean zero and covariance matrix \(B^{-1} \Sigma [B^{-1}]^T\).

One may notice that Theorem 1 is an extension to the multivariate case from Theorem 4 discussed in Carroll et al. (1997) concerning univariate data. This extension is readily established by adapting the arguments of Carroll et al. (1997). As association parameters \(\phi\) is also of prime interest here, it is the major concern to establish the asymptotic distribution of the estimator \(\hat{\phi}\) along with that of \((\hat{\alpha}, \hat{\beta})\). Let \(J = E(U_{i\phi} U_{i\phi}^T)\) and \(H = E[(-\partial / \partial \phi^T) U_{i\phi}]\). If \(\alpha, \beta\) and \(\theta(.)\) are all known, estimating functions \(U_{i\phi}\) are regular parametric unbiased estimating functions of
\( \phi \), and thereby it is straightforward to establish that \( \sqrt{n}(\hat{\phi} - \phi) \sim N(0, H^{-1}J[H^{-1}]^\top) \), according to Liang and Zeger (1986). When \( \alpha, \beta \) and \( \theta(.) \) are unspecified and estimated, variation in the estimators \( \hat{\alpha}, \hat{\beta} \) and \( \hat{\theta}(u) \) must be taken into account. If \( \theta(.) \) is a known parametric function, one may easily adapt the arguments in Yi and Cook (2002) to work out the asymptotic distribution of \( \sqrt{n}(\hat{\phi} - \phi) \). However, here \( \theta(.) \) is unknown and it is estimated locally, we need to incorporate this local estimation variability into the asymptotic variance of \( \sqrt{n}(\hat{\phi} - \phi) \) as well. This unknown \( \theta(.) \) function presents a challenge in establishing the asymptotic distribution for the estimators, and it is this feature that distinguishes the current work from the existing results. Assuming the fourth moment of \( Y_i \) exists, we establish the joint asymptotic distribution of the estimator \( (\hat{\alpha}, \hat{\beta}, \hat{\phi}) \) in Theorem 2. The proof is given in Appendices 2 and 3.

**Theorem 4.2** Under the conditions in Appendix 1, as \( nh^4 \to 0 \) and \( nh^2 / \log(1/h) \to \infty \),
\[
\sqrt{n}\{\begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \\ \hat{\phi} - \phi \end{pmatrix}^\top \} \text{ has the limiting distribution } N(0, B^*\Sigma^*[B^*]^{-1}^\top).
\]

Here \( B^* \) and \( \Sigma^* \) are defined analogously to \( B \) and \( \Sigma \). Their detailed expressions are presented in Appendix 2. Inferences about parameters \( \alpha, \beta \) and \( \phi \) may, using Theorem 2, be conducted by replacing \( B^* \) and \( \Sigma^* \) with their empirical estimates. As implementation of these variance estimates is computationally difficult, in practice, we often employ bootstrap methods for variance estimates when actually conducting inference. See Lin and Carroll (2000) and Liang, Wang, Robins, and Carroll (2004), for instance.

### 4.2 Bandwidth Selection

In the implementation of the procedures above, choosing an appropriate bandwidth \( h \) is very crucial. As bandwidth \( h \) affects both bias and variance estimate, there is a trade-off between suitable bias and variance estimate. Bias correction requires the choice of a relatively small bandwidth, whereas variance estimate needs to choose a large value of bandwidth. In principle, bandwidth selection is data driven, and traditional methods such as cross-validation approach may be applied to select a proper bandwidth \( h \) based on available data. However, as pointed out in Fan, Heckman, and Wand (1995), this approach could perform poorly in some settings with a large magnitude of sample variation produced, hence it is not regarded as a sensible bandwidth selection rule for practical use. Instead, “plugging in” method may be a promising candidate for bandwidth selection.
Fan et al. (1995) discussed this approach to handle local polynomial regression under the framework of generalized linear models. Ruppert, Sheather, and Wand (1995) explored this method of bandwidth selection with local least squares regression.

Along the same line Carroll et al. (1997) presented a detailed discussion on bandwidth selection for generalized partially linear single index models. In particular, to estimate $\alpha$ and $\beta$ at rate $n^{-1/2}$, one needs to undersmooth the nonparametric function $\theta(\cdot)$, which is a standard routine in the kernel literature (e.g., Carroll et al. 1997). We note that the usual optimal bandwidth rate for nonparametric regression, $\hat{h}_{opt}$ of order $n^{-1/5}$, does not satisfy the condition in Theorems 4.1 and 4.2. A sensible choice of bandwidth $h$ is generally difficult. Carroll et al. (1997) suggested an ad hoc bandwidth, given by $\hat{h}_{opt} \times n^{-2/15} = O(n^{-1/5}) \times n^{-2/15} = O(n^{-1/3})$, which satisfies the requirements $nh^4 \to 0$ and $nh^2/(\log n)^2 \to \infty$ in Theorems 4.1 and 4.2.

5 Numerical Studies

5.1 An Example

We apply the proposed methods to analyze a subset of Genetic Analysis Workshops (GAW) 13 data arising from Cohort 2 of the Framingham Heart Study. The Framingham Heart Study is an ongoing prospective study of risk factors for cardiovascular disease (CVD). The objective of the Framingham Heart Study was to identify common factors or characteristics that contribute to CVD by following its development over a long period of time in a large group of participants who had not yet developed overt symptoms of CVD or suffered a heart attack or stroke.

In the analysis here we consider the data set consisting of 203 families each having 4 members with the baseline measurements. High blood pressure is an important risk factor for cardiovascular disease and is a leading cause of mortality in industrialized countries. As high blood pressure is a complex disorder that results from environmental and genetic factors and their interactions, and other study indicates that blood pressure increases with age (Kraft, Bauman, Yuan, and Horvath 2003). It is of interest to study how blood pressure is influenced by the risk factors and how individuals within the same family may be associated. The covariates of interest include age, gender, high density lipoprotein (HDL) and body mass index (BMI) $(\text{BMI}=\text{weight (kg)}/\text{height}^2)$.
\( m^2 \)). Let \( Y_{ij} = 1 \) if subject \( j \) in family \( i \) has high blood pressure, and \( Y_{ij} = 0 \) otherwise.

We consider a semiparametric regression model for the mean response which is specified as

\[
\logit \mu_{ij} = \beta x_{ij} + \theta (\alpha_1 z_{ij1} + \alpha_2 z_{ij2} + \alpha_3 z_{ij3}),
\]

where \( x_{ij} \) is gender, taking value 1 for male and 0 otherwise, \( z_{ij1} \) is age, \( z_{ij2} \) is HDL, and \( z_{ij3} \) is BMI. \( z_{ij1}, z_{ij2} \) and \( z_{ij3} \) are standardized as \( (z_{ijr} - \bar{z}_{..r})/s_{..r} \), where \( \bar{z}_{..r} \) and \( s_{..r} \) represent the sample mean and standard deviation of \( z_{ijr} \)'s, respectively, \( r = 1, 2, 3 \). Exchangeable association structure is modeled here with \( \log \psi_{ijk} = \phi \), for \( j \neq k \).

In Table 1 we report the parameter estimates, standard errors and p-values. The analysis results suggest that gender has an effect on having high blood pressure. Males are more likely to have high blood pressure than females. At significance level 0.05, the analysis suggests that age plays an important role in predicting high blood pressure. As people get older, they are more prone to have higher blood pressure. There is evidence to support a moderate positive effect of HDL on having high blood pressure. Strong evidence indicates that BMI has a statistically significant impact on high blood pressure. An individual with a larger BMI has a larger chance to have higher blood pressure. It is noted that there is no evidence to support an existing association among response measurements of family members.

*Insert Table 1 about here*

To understand if there is a curvature relationship between the response and covariate variables, we plot \( \logit \mu_{ij} \) against the single index \( \theta(\cdot) \) respectively for females and males data. The patterns are displayed in Figure 1. It is seen that there are nonlinear trends for both data sets. Thereby using a nonlinear term in the regression (5.6) for \( \logit \mu_{ij} \) is perhaps very plausible.

### 5.2 Simulation Study

We conduct a simulation study to evaluate the performance of the proposed methods. Here we focus on pairwise association with higher order association being constrained as 0. That is, generate binary vector \( y_i = (y_{i1}, y_{i2}, \ldots, y_{im})^T \) from the joint density function

\[
f(y_{i1}, y_{i2}, \ldots, y_{im}) = \prod_{j=1}^{m} \mu_{ij}^{y_{ij}} (1 - \mu_{ij})^{1-y_{ij}}
\]

(5.7)
\[ 1 + \sum_{j<k} \rho_{i;jk} \frac{y_{ij} - \mu_{ij}}{\sqrt{v_{i;jj}}} \cdot \frac{y_{ik} - \mu_{ik}}{\sqrt{v_{i;kk}}} \]

where \( \rho_{i;jk} \) is the correlation coefficient of \( Y_{ij} \) and \( Y_{ik} \), given by
\[ \rho_{i;jk} = \frac{(\mu_{i;jk} - \mu_{ij}\mu_{ik})}{\sqrt{v_{i;jj}v_{i;kk}}} \]

The mean responses are modeled as
\[ \text{logit } \mu_{ij} = \beta x_{ij} + \theta(\alpha_1 z_{i1} + \alpha_2 z_{i2} + \alpha_3 z_{i3}) \]

where we take \( \theta(t) = \sin[\pi(t - 1.355\sqrt{3}/6)/(1.645\sqrt{3}/3)] \) as the same form in Carroll et al. (1997). We consider an exchangeable association structure by specifying
\[ \log \psi_{i;jk} = \phi \]

Covariates \( x_{ij} \) are generated according to the binomial distribution \( Bin(1, 0.5) \) and covariates \( z_{ij} \) are generated from the uniform distribution \( U[0, 1] \). Set \( \beta = 0.3 \) and \( \alpha_1 = \alpha_2 = \alpha_3 = 1/\sqrt{3} \).

Various configurations of \( \phi \) are considered to reflect different strengths of association. In the simulation we focus on a setting where \( m = 4 \) and \( n = 100 \). 100 simulations are run for each parameter configuration. Table 2 contains the differences (Bias) between the true values and the estimates and the empirical standard errors (S.E.) for the regression and association parameters.

Insert Table 2 about here

It can be seen that the estimates for the linear coefficient \( \beta \) are relatively stable as the odds ratio \( \psi_{i;jk} \) varies. The estimates of the nonlinear coefficients \( \alpha \) seem more affected by different strengths of the association. The finite sample biases of the estimates for both \( \beta \) and \( \alpha \) are reasonably small in a context of semiparametric regression. It seems that the estimates of the association parameter \( \phi \) tend to be more off the true value as the association becomes stronger. However, the estimates are still reasonably in good agreement with the true values. We notice that \( \hat{\phi} \) has the smallest standard errors among all the estimators. It is not surprising that the empirical standard errors for the estimates of the linear coefficient \( \beta \) are smaller than those of \( \hat{\alpha} \). The empirical standard errors for all the estimates do not appear to vary much as the strength of the association becomes different. This simulation demonstrates that the proposed method gives rise to reasonable estimates for both the mean and association parameters.
5.3 Model Misspecification: Case 1

When analyzing a real data set, it is advisable to check the data first to see if there are any noticeable nonlinear patterns. Model misspecification may, in principle, lead to erroneous or misleading inference results. Specifically, there are two typical cases of model misspecification. In this subsection we consider the first case of model misspecification. Namely, if the true underlying model follows (2.1) but we adopt a standard logistic regression model, then the resultant estimator could be biased. Here we conduct a simulation to numerically evaluate the bias induced using such a misspecified model. We simulate the responses from the marginal model (2.1) together with (5.9) as in Section 5.1, but we fit the simulated data to a usual logistic regression model given by

$$\text{logit } \mu_{ij} = \beta x_{ij} + \alpha_1 z_{i1} + \alpha_2 z_{i2} + \alpha_3 z_{i3}$$  (5.10)

in combination with (5.9). The parameter configurations and the number of simulations are set as the same as in Section 5.1.

Table 3 reports on the results of the empirical biases and standard errors. It can be seen that the estimates for the mean parameters $\beta$ involve larger biases, in contrast to the results in Table 2 that are produced by using the true model. The biases contained in the estimates of the nonlinear parameters $\alpha$ are more pronounced than those in the linear parameters $\beta$. It is not surprising that the estimates for the association parameter $\phi$ do not contain remarkable bias, indicating that the estimator $\hat{\phi}$ is fairly reliable for estimating parameter $\phi$. This may be partly explained by that the estimating function $U_{i\phi}$ for parameter $\phi$ is still unbiased even if the marginal mean model is misspecified, and biased estimates in $\beta$ and $\alpha$ do not radically impact the estimating function $U_{i\phi}$ when proceeding with estimation of the association parameter $\phi$. One may also notice that the empirical standard errors for $\hat{\phi}$ are the smallest in contrast to those of $\hat{\beta}$ and $\hat{\alpha}$.

Insert Table 3 about here

5.4 Model Misspecification: Case 2

Now we discuss another case of model misspecification. That is, we consider a scenario when a standard logistic regression model well fits the data, but we have used model (2.1) to perform
estimation. It is therefore interesting to understand how the proposed method may perform in terms of the change in bias and efficiency. As a linear function is a special form of a nonlinear function, it is expected that the resulting estimator is still consistent, but there may be a possible efficiency loss incurred. We generate the response measurements from the marginal model given by (5.10) together with the association model (5.9). We fit models (5.10) and (5.9) to the simulated data, and this is called Method 1. Method 2 is to fit the simulated data with models (5.8) and (5.9). Again the same simulation settings as in Section 5.1 are used here.

Table 4 presents the empirical biases and standard errors for the estimates of both the mean and association parameters which are obtained from both methods. The biases for the estimates of $\beta$ are reasonably comparable from both Method 1 and Method 2 and the magnitudes are fairly small. This confirms that estimator $\hat{\beta}$ obtained from Method 2 is consistent for parameter $\beta$, just like that obtained from Method 1. It can be seen that Method 2 produces more obvious bias in the estimates for $\alpha$ than that of Method 1, yet the magnitudes seem acceptable relative to those reported in Table 2. For the association parameter $\phi$ the estimates obtained from both methods are in good agreement, and the resulting biases are in a reasonable small range. We may notice that Method 2 tends to produce larger empirical standard errors for $\hat{\beta}$ and $\hat{\phi}$ than Method 1 does, but the difference does not seem considerable. In other words, Method 2 may incur somewhat efficiency loss in estimating the linear coefficient $\beta$ and association parameter $\phi$, but the effect is not profound. However, such a trend is not observed for the estimates for $\alpha$.

Insert Table 4 about here

6 Discussion

In this paper we describe a semiparametric approach to analyze clustered binary data. Interest here lies in estimation of the association coefficients in addition to the marginal mean parameters. Specifically, we model both the mean response and the association using semiparametric and parametric regressions, respectively, and this allows us to relate the response with covariates by facilitating richer structures of both mean response and association. The simulation studies demonstrate that the proposed methods work well under various situations. More generally, specification of
association may be based on a semiparametric form like

\[ \log \psi_{i,j,k} = u_{i,j,k}^T \phi + \eta^*(w_{i,j,k}^T \delta^*) \]  

(6.11)

where \( \eta^*(.) \) is an unknown smoothing function. Extension to including (6.11) is straightforward by adapting the arguments in the paper, though a more complicated presentation is needed.

The methods we describe here have applications in a wide variety of settings. They can also be generalized to accommodating data with more complex association structures. In many situations, clustered data may arise from longitudinal studies. Clustered longitudinal data feature both a cross-sectional and a longitudinal correlation structure and interest often lies in the strengths of both types of association (Yi and Cook 2002). The proposed methods may be adapted to handle longitudinal data arising in clusters.

We note that in the estimation algorithm the working independence matrix is employed in stage 1 when conducting estimation of the mean parameters. This is basically motivated by the findings in Lin and Carroll (2000, 2001a). It is interesting to modify the current development along the lines of Wang (2003) and Wang, Carroll, and Lin (2005) to incorporate the true correlation structure in both stage 1 and stage 2 of the estimation procedures. Efficiency comparison would thereby be of interest.

Acknowledgements

The research of Yi and He was supported by the Natural Sciences and Engineering Research Council of Canada. The research of Liang was partially supported by the NIAD/NIH grants AI62247, AI059773, and AI50020. The authors thank Boston University and NHLBI for providing the data set from the Framingham Heart Study (No. N01-HC-25195) in the illustration. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University. This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI.
Appendix 1: Conditions

Without exception detailed technical conditions are needed here to guarantee rigorous proofs. Below we just outline several key assumptions with the detailed list of conditions omitted. For more details see Carroll et al. (1997).

(a) The density function of $z_{ij}, f(z)$, has a continuous second derivative on its support.
(b) The density function of $z_{ij}^T \alpha, k(t)$, is positive and uniformly continuous for $\alpha$ in a neighborhood of its true value.
(c) $\theta''(u)$ is continuous on its support.
(d) The random vector $x_{ij}$ is assumed to have a bounded support.
(e) $K(\cdot)$ is a symmetric probability density function with bounded support.

In the following development the identities are valid to the order of $O_p(h^2) + o_p(n^{-1/2})$.

Appendix 2: Several Lemmas

For the establishment of Theorem 2 we need to introduce some notation that corresponds to the form of $\epsilon_i$ in estimating functions $U_{i\phi}(\alpha, \beta, \theta(\cdot), \phi)$. Note that $\epsilon_i$ may be written as $y_i^* - \xi_i$ where $y_i^* = (1_{m-1}^T y_{i1}, 1_{m-2}^T y_{i2}, \ldots, 1_{i-1}^T y_{i,m-1})^T$. In the sequel superscript $*$ is used to indicate the ordering similar to that in $y_i^*$, and double superscript $**$ is for the one similar to that in $\xi_i$. To be more specific, let $\theta^*(z, \alpha) = (1_{m-1}^T \theta(z_{i1}^T \alpha), 1_{m-2}^T \theta(z_{i2}^T \alpha), \ldots, 1_{i-1}^T \theta(z_{i,m-1}^T \alpha))^T$ and $\theta^{**}(z, \alpha) = (\theta(z_{i1}^T \alpha), \theta(z_{i2}^T \alpha), \ldots, \theta(z_{i,m}^T \alpha), \theta(z_{i,m+1}^T \alpha), \ldots, \theta(z_{i,m+n}^T \alpha))^T$ be $m* \times 1$ vectors, where $m* = m(m-1)/2$. Similar notation is defined for $z_i^*$ and $z_i^{**}$.

Let $Q_{x}(z_i, \alpha) = ([1_{m-1} Q_z(z_{i1}^T \alpha)]^T, [1_{m-2} Q_z(z_{i2}^T \alpha)]^T, \ldots, [1_{i-1} Q_z(z_{i,m-1}^T \alpha)]^T)^T$ be an $m* \times p$ matrix. Let $Q_x(s)(z_i, \alpha) = ([Q_x(z_{i,s}^T \alpha)]^T, [Q_x(z_{i,s+1}^T \alpha)]^T, \ldots, [Q_x(z_{i,m}^T \alpha)]^T)^T$ be an $(m-s+1) \times p$ matrix for $s = 1, \ldots, m$, and $Q_{x}^{**}(z_i, \alpha) = ([Q_x(2)(z_i, \alpha)]^T, \ldots, [Q_x(m)(z_i, \alpha)]^T)^T$ be an $m* \times p$ matrix. $Q_{x}^*(z_i, \alpha)$, and $Q_{x}^{**}(z_i, \alpha)$ are defined analogously.

Let $\theta_{\text{diag}}''(z_i, \alpha)$ and $\theta_{\text{diag}}^{**}(z_i, \alpha)$ be diagonal block matrices whose $s$th diagonal blocks are $\theta'(z_{is}^T \alpha) I_{(m-s)(m-s)}$ and $\text{diag}(\theta'(z_{is,s+1}^T \alpha), \ldots, \theta'(z_{im}^T \alpha))$, respectively, $s = 1, \ldots, m-1$. Here $I_{r \times r}$ denotes the $r \times r$ identity matrix. Similar definition applies to $Q_{x}^{*1}(z_i, \alpha)$ and $Q_{x}^{**1}(z_i, \alpha)$, where
their $s$th diagonal blocks are $Q_{11}(z^s_{i,s} \alpha)I_{m-s} \times (m-s)$ and \text{diag}(Q_{11}(z^s_{i,s+1} \alpha), ..., Q_{11}(z^s_{i,m} \alpha))$, respectively, $s = 1, ..., m - 1$. Let $\xi^s_{\text{diag}}(z_i, \alpha)$ and $\xi^s_{\text{diag}}(z_i, \alpha)$ be diagonal block matrices whose $s$th diagonal blocks are $\text{diag}\{\frac{\partial \xi_{s,j}}{\partial (z^s_{i,j} \alpha)}, j = s + 1, ..., m\}$ and $\text{diag}\{\frac{\partial \xi_{s,j}}{\partial (z^s_{i,j} \alpha)}, j = s + 1, ..., m\}$, $s = 1, ..., m - 1$.

Define $A^s_i = \left(\frac{\partial \xi^s}{\partial \phi}\right)W_i^{-1}\xi^s_{\text{diag}}(z_i, \alpha)Q_{11}^{-1}(z_i, \alpha)$, and $A^{**}_{11}$ being $A^s_i$ but replacement of $\xi^s_{\text{diag}}$ and $Q_{11}$ by $\xi^{**}_{\text{diag}}$ and $Q_{11}^{**}$: Let $P^*_i = \{E(A^s_i 1_{m*}|U_{k1}), ..., E(A^{**}_i 1_{m*}|U_{k1})\}$. Denote $E(A^s_i 1_{m*}|U_{km})\Delta_i$, $P^{**}_i = \{E(A^s_i 1_{m*}|U_{k1}), ..., E(A^{**}_i 1_{m*}|U_{km})\} \Delta_i$. Denote

$$B^*_i = \left(\frac{\partial \xi^s}{\partial \phi}\right)W_i^{-1}\left(\frac{\partial \xi^s}{\partial \phi}\right) - \{\xi^s_{\text{diag}}(z_i, \alpha)Q^s(z_i, \alpha) + A^s_i Q^s_{z}Q^s_{z}(z_i, \alpha)\}$$

with $B^{**}_{ij} = \left(\frac{\partial \xi^s}{\partial \phi}\right)W_i^{-1}\left(\xi^s_{\text{diag}}(z_i, \alpha)\theta^s_{\text{diag}}(z_i, \alpha)z_i^* + \xi^{**}_{\text{diag}}(z_i, \alpha)\theta^{**}_{\text{diag}}(z_i, \alpha)z_i^{**}\right) - \{A^s_i Q^s_{z}(z_i, \alpha) + A^{**}_{i} Q^{**}_{z}(z_i, \alpha)\}$, and $B^* = \begin{pmatrix} \tilde{B}_i \\ B^*_i \end{pmatrix}$, where $\tilde{B}_i = (B_i 0)$ is a $(q + p + a) \times (q + p + a)$ matrix. Here $a$ is the dimension of $\phi$. Let $Q_i = \begin{pmatrix} P_i & \widetilde{\Delta}_i \\ - (P^*_i + P^{**}_i) \end{pmatrix}$

$$W_i^{-1} - \left(\frac{\partial \xi^s}{\partial \phi}\right)W_i^{-1}\left(\frac{\partial \xi^s}{\partial \phi}\right) \begin{pmatrix} y_i - \mu_i \\ y_i^* - \xi_i \end{pmatrix} \text{cov} \begin{pmatrix} y_i - \mu_i \\ y_i^* - \xi_i \end{pmatrix}|U_i]$$

Assuming the fourth moment of $Y_i$ exists, we denote $\Omega_i = \text{COV} \begin{pmatrix} y_i - \mu_i \\ y_i^* - \xi_i \end{pmatrix}|U_i]$ and $\Sigma^* = E(Q_i \Omega_i Q_i^t)$.

Further we introduce additional notation for the proof of Theorem 2. Let $\hat{\theta}(z_i, \alpha; \bullet) = \{(1_{m-1} \hat{\theta}(z_{i1}^* \alpha; \bullet))^{\top}, (1_{m-2} \hat{\theta}(z_{i2}^* \alpha; \bullet))^{\top}, ..., (1_{i,m-1} \hat{\theta}(z_{im-1}^* \alpha; \bullet))^{\top}\}^{\top}, \hat{\theta}^{**}(z_i, \alpha; \bullet) = \{(z_{i1} \alpha; \bullet), \hat{\theta}(z_{i2} \alpha; \bullet), \hat{\theta}(z_{i3} \alpha; \bullet), ..., \hat{\theta}(z_{im} \alpha; \bullet), \hat{\theta}(z_{i,m} \alpha; \bullet)\}^{\top}$, where $\bullet$ denotes $(\tilde{\alpha}, \tilde{\beta})$. Let $e_i = \hat{W}_i^{-1}(y_i - \mu_i)$, $e^*_i = W_i^{-1}(y_i^* - \xi_i)$, and $d_k(u) = 1_m \Delta_k \bar{K}_{kh}(u)$. Define

$$D_k = \{[d_k(z_{i1}^* \alpha)]^\top, [d_k(z_{i2}^* \alpha)]^\top, ..., [d_k(z_{im}^* \alpha)]^\top\}^\top,$$

$$D^*_k = \{[1_{m-1}d_k(z_{i1}^* \alpha)]^\top, [1_{m-2}d_k(z_{i2}^* \alpha)]^\top, ..., [1_{i,m-1}d_k(z_{im-1}^* \alpha)]^\top\}^\top,$$

$$D^{**}_k = \{[d_k(z_{i2}^* \alpha)]^\top, [d_k(z_{i3}^* \alpha)]^\top, ..., [d_k(z_{i,m}^* \alpha)]^\top\}^\top,$$

$$\ldots [d_k(z_{i,m}^* \alpha)]^\top, ..., [d_k(z_{i,m}^* \alpha)]^\top\}^\top.$$
Let
\[
Q_n(u) = \frac{1}{n} \sum_{i=1}^{n} \Gamma_i(u) \Delta_i K_i(u) \bar{W}_i^{-1} \Delta_i \Gamma_i(u),
\]
\[
Q_{nx}(u) = \frac{1}{n} \sum_{i=1}^{n} \Gamma_i(u) \Delta_i K_i(u) \bar{W}_i^{-1} \Delta_i x_i,
\]
\[
Q_{nz}(u) = \frac{1}{n} \sum_{i=1}^{n} \Gamma_i(u) \Delta_i K_i(u) \bar{W}_i^{-1} \Delta_i \theta_{\text{diag}}(z_i, \alpha) z_i.
\]

Denote \(Q_{nst}(u)\) as the \((s, t)\) element of \(Q_n(u)\), \(s, t = 1, 2\). Let \(Q_{nx1}(u)\) and \(Q_{nz1}(u)\) be the 1st row of \(Q_{nx}(u)\) and \(Q_{nz}(u)\), respectively.

**Lemma 1:** If the conditions of Theorem 1 hold, then, as \(n \to \infty\), \(Q_{n11}(u) \to Q_{11}(u), Q_{n12}(u) \to 0\), \(Q_{n21}(u) \to 0\), hence \([Q_n(u)]^{-1} \to \begin{pmatrix} Q_{11}^{-1}(u) & 0 \\ 0 & Q_{22}^{-1}(u) \end{pmatrix}\) for some function \(Q_{22}(u)\), and \(Q_{nx1}(u) \to Q_x(u), Q_{nx1}(u) \to Q_z(u)\) in probability.

**Proof:** Note that \(1_m^T \bar{\Delta} \bar{K} h(u) = (\tilde{\mu}_{11}^{(1)} K_h(z_i^T \tilde{\alpha} - u), \ldots, \tilde{\mu}_{im}^{(1)} K_h(z_i^T \tilde{\alpha} - u))\), and \(\bar{W}_i^{-1} \Delta_i 1_m = \text{diag}(w_i^{s,s}, s = 1, \ldots, m) \cdot (\tilde{\mu}_{i1}^{(1)}, \ldots, \tilde{\mu}_{im}^{(1)})^T\). Then
\[
Q_{n11}(u) = \sum_{s=1}^{m} \left\{ \frac{1}{n} \sum_{i=1}^{n} [\tilde{\mu}_{is}^{(1)}]^2 K_h(z_i^T \tilde{\alpha} - u) w_i^{s,s} \right\}
\]
and this converges in probability to \(\sum_{s=1}^{m} E[(\tilde{\mu}_{is}^{(1)})^2 w_1^{s,s}|U_1 = u]\). Other convergences may be shown similarly.

**Lemma 2:**
\[
\hat{\theta}^*(z_i, \alpha; \bullet) - \theta^*(z_i, \alpha) = Q_{x11}^{*-1}(z_i, \alpha) \left\{ \frac{1}{n} \sum_{k=1}^{n} D_{ik}^T e_k - Q_{x}^*(z_i, \alpha)(\hat{\beta} - \beta) \right\} - Q_{x}^*(z_i, \alpha)(\hat{\alpha} - \alpha).
\]
\[
\hat{\theta}^{**}(z_i, \alpha; \bullet) - \theta^{**}(z_i, \alpha) = Q_{x11}^{**-1}(z_i, \alpha) \left\{ \frac{1}{n} \sum_{k=1}^{n} D_{ik}^{**} e_k - Q_{x}^{**}(z_i, \alpha)(\hat{\beta} - \beta) \right\} - Q_{x}^{**}(z_i, \alpha)(\hat{\alpha} - \alpha).
\]

**Proof:** It suffices to show elementwisely that for \(j = 1, \ldots, m\),
\[
\hat{\theta}(z_{ij}^T \alpha; \bullet) - \theta(z_{ij}^T \alpha) = Q_{11}^{-1}(z_{ij}^T \alpha) \left\{ \frac{1}{n} \sum_{k=1}^{n} d_{ik}(z_{ij}^T \alpha)e_k - Q_{x}(z_{ij}^T \alpha)(\hat{\beta} - \beta) \right\} - Q_{x}(z_{ij}^T \alpha)(\hat{\alpha} - \alpha).
\]
It follows from (3.2) that
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma_i^r(u) \Delta_i \tilde{K}_{ih}(u) e_i - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma_i^r(u) \Delta_i \tilde{K}_{ih}(u) \tilde{W}_i^{-1} (\tilde{\mu}_i - \mu_i) = 0. \] (6.13)

By the Taylor series expansion, the second term of (6.13) can be decomposed as
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma_i^r(u) \Delta_i \tilde{K}_{ih}(u) \tilde{W}_i^{-1} \left( \Delta_i \Gamma_i^r(u) (\tilde{a} - a) + \Delta_i [\tilde{\beta} - \beta] + \theta_d'(z_i, \alpha) z_i (\tilde{\alpha} - \alpha) \right) \]

at convergence of \((\tilde{\beta}, \tilde{\alpha})\). This expression and (6.13) imply that
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma_i^r(u) \Delta_i \tilde{K}_{ih}(u) \tilde{W}_i^{-1} \Delta_i \Gamma_i^r(u) \left( \tilde{a}_0 - a_0 \right) \]
\[ = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma_i^r(u) \Delta_i \tilde{K}_{ih}(u) e_i - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma_i^r(u) \Delta_i \tilde{K}_{ih}(u) \tilde{W}_i^{-1} \Delta_i x_i (\tilde{\beta} - \beta) \]
\[ - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma_i^r(u) \Delta_i \tilde{K}_{ih}(u) \tilde{W}_i^{-1} \Delta_i \theta_d'(z_i, \alpha) z_i (\tilde{\alpha} - \alpha). \] (6.14)

Then by Lemma 1, we obtain, as \(n \to \infty\),
\[ \tilde{a}_0(u) - a_0(u) = Q_{11}^{-1}(u) \left\{ \frac{1}{n} \sum_{i=1}^{n} \Gamma_i^r \Delta_i \tilde{K}_{ih}(u) e_i \right. \]
\[ \left. - Q_e(u) (\tilde{\beta} - \beta) - Q_e(u) (\tilde{\alpha} - \alpha) \right\}. \] (6.15)

Therefore, letting \(u = z_i^2 \alpha\) results in (6.12).

**Lemma 3:**
\[ \sum_{i=1}^{n} A_i^* \frac{1}{n} \sum_{k=1}^{n} D_{ik} e_k = \sum_{k=1}^{n} \left( E(A_k^{*1} 1_{m^*} | U_{k1}), ..., E(A_k^{*1} 1_{m^*} | U_{km}) \right) \tilde{\Delta}_k e_k, \]
\[ \sum_{i=1}^{n} A_i^{**} \frac{1}{n} \sum_{k=1}^{n} D_{ik} e_k = \sum_{k=1}^{n} \left( E(A_k^{**1} 1_{m^*} | U_{k1}), ..., E(A_k^{**1} 1_{m^*} | U_{km}) \right) \tilde{\Delta}_k e_k, \]
\[ \sum_{i=1}^{n} A_i^* \frac{1}{n} \sum_{k=1}^{n} D_{ik} e_k = \sum_{k=1}^{n} \left( E(A_k 1_{m} | U_{k1}), ..., E(A_k 1_{m} | U_{km}) \right) \tilde{\Delta}_k e_k, \]

where \(\tilde{\Delta}_i\) is \(\Delta_i\) with \(\tilde{\alpha}\) and \(\tilde{\beta}\) respectively replaced by the true parameters \(\alpha\) and \(\beta\).

**Proof:** Write \(A_i^* = (A_{m-1}, A_{m-2}, ..., A_{1}^*)\) where \(A_{r,i}^*\) is the corresponding submatrix of \(A_i^*\) of \(r\) columns. As \(\sum_{i=1}^{n} A_i^* \frac{1}{n} \sum_{k=1}^{n} D_{ik} e_k = \sum_{k=1}^{n} (\frac{1}{n} \sum_{i=1}^{n} A_i^* D_{ik}) e_k\), and \(d_k(u) = (\tilde{\mu}_k^{(1)} K_k(\tilde{U}_{k1} - \)
\[
\frac{1}{n} \sum_{i=1}^{n} A_i^* D_{ik} = \frac{1}{n} \sum_{i=1}^{n} (A_{m-1,i}^* A_{m-2,i}^* \ldots A_{i,i}^*) \begin{pmatrix}
1_{m-1} d_k(z_{i1}^* \alpha) \\
1_{m-2} d_k(z_{i2}^* \alpha) \\
\vdots \\
1_1 d_k(z_{im}^T \alpha)
\end{pmatrix}
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{r=1}^{m-1} \{ A_{m-r,i}^* 1_{m-r} \mu_{k1} K_h(\hat{U}_{k1} - z_{ir}^* \alpha), \ldots, A_{m-r,i}^* 1_{m-r} \mu_{km} K_h(\hat{U}_{km} - z_{ir}^* \alpha) \}.
\]

Now consider the \( j \)th component of this expression \( j = 1, 2, \ldots, m \), then we have
\[
\frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{r=1}^{m-1} A_{m-r,i}^* 1_{m-r} \mu_{k1} K_h(\hat{U}_{k1} - z_{ir}^* \alpha) \right\} \rightarrow_p \sum_{r=1}^{m-1} E(A_{m-r,k}^* U_{kj}) = u_{kj} 1_{m-r} \mu_{k1} K_h(\hat{U}_{km} - z_{ir}^* \alpha),
\]
where \( \mu_{kj} \) is \( \mu_{k1} \) with \( \hat{\alpha} \) and \( \hat{\beta} \) respectively replaced by the true parameters \( \alpha \) and \( \beta \). Hence,
\[
\frac{1}{n} \sum_{i=1}^{n} A_i^* D_{ik} \rightarrow_p \{ E(A_{k1}^* 1_{m*} U_{k1}) \ldots E(A_{km}^* 1_{m*} U_{km}) \} \Delta_k
\]
since \( A_k^* 1_{m*} = \sum_{r=1}^{m-1} A_{m-r,k}^* 1_{m-r} \). The other expressions can be proved similarly.

Modifying the arguments in Carroll et al. (1997), we may prove the following lemma.

**Lemma 4:**
\[
\hat{\theta}^*(z_i, \hat{\alpha}; \bullet) - \theta^*(z_i, \alpha) = \theta_{\text{diag}}^*(z_i, \alpha) z_i^* (\hat{\alpha} - \alpha) + (\hat{\theta}^*(z_i, \alpha; \bullet) - \theta^*(z_i, \alpha)),
\]
\[
\hat{\theta}^{**}(z_i, \hat{\alpha}; \bullet) - \theta^{**}(z_i, \alpha) = \theta_{\text{diag}}^{**}(z_i, \alpha) z_i^{**} (\hat{\alpha} - \alpha) + (\hat{\theta}^{**}(z_i, \alpha; \bullet) - \theta^{**}(z_i, \alpha)),
\]
\[
\xi_i - \hat{\xi}_i(\hat{\alpha}, \hat{\beta}) = \frac{\partial \xi_i}{\partial \beta} (\hat{\beta} - \beta) - \frac{\partial \xi_i}{\partial \phi} (\hat{\phi} - \phi) - \xi_{\text{diag}}(z_i, \alpha) (\hat{\theta}^*(z_i, \hat{\alpha}; \bullet) - \theta^*(z_i, \alpha))
\]
\[
\quad - \xi_{\text{diag}}(z_i, \alpha) (\hat{\theta}^{**}(z_i, \hat{\alpha}; \bullet) - \theta^{**}(z_i, \alpha)).
\]

**Appendix 3: Proof of Theorem 2**

The proof comprises the following three steps.

1. Adapting the proof of Theorem 4 in Carroll et al. (1997) under the assumption \( nh^2 / \log(1/h) \rightarrow \infty \), we have
\[
\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix}
z_i^T \hat{\theta}_{\text{diag}}^*(z_i, \hat{\alpha}; \bullet) \\
x_i^T \Delta_i W_i^{-1} \{ \mu_i - \hat{\mu}_i(\hat{\alpha}, \hat{\beta}) \}
\end{pmatrix}
\]

19
\[
\sum_{i=1}^{n} \left\{ \left( \frac{\partial \xi_i^T}{\partial \phi} \right) W_i^{s-1}(\xi_i - \hat{\xi}_i(\alpha, \beta, \phi)) + \xi_{\text{diag}}(z_i, \alpha)(\hat{\theta}^{**}(z_i, \alpha; \bullet) - \theta^{**}(z_i, \alpha)) \right\} = \mathcal{J}.
\]

Using Lemma 4, we have
\[
\mathcal{J} = -\frac{1}{n} \sum_{i=1}^{n} B_i \sum_{i=1}^{n} A_i \sum_{k=1}^{n} D_{ik} e_k.
\]

A direct but tedious derivation yields
\[
\mathcal{J} = -B^*_i \left\{ \left( \alpha - \hat{\alpha} \right)^T, \left( \beta - \hat{\beta} \right)^T, (\hat{\phi} - \phi)^T \right\}^T - \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\partial \xi_i^T}{\partial \phi} \right) W_i^{s-1} \left\{ \xi_{\text{diag}}(z_i, \alpha) Q_{11}^{-1}(z_i, \alpha) \frac{1}{n} \sum_{k=1}^{n} D_{ik} e_k \right. \\
+ \xi_{\text{diag}}^*(z_i, \alpha) Q_{11}^{-1}(z_i, \alpha) \frac{1}{n} \sum_{k=1}^{n} D_{ik} e_k \left. \right\} \\
= -B^*_i \left\{ \left( \alpha - \hat{\alpha} \right)^T, \left( \beta - \hat{\beta} \right)^T, (\hat{\phi} - \phi)^T \right\}^T - \frac{1}{n} \sum_{i=1}^{n} \left\{ A_i^* D_{ik} + A_i^{**} D_{ik}^* \right\} e_k.
\]

3. Now as \( y_i^* = \epsilon_i + \xi_i \), working on (3.3), (3.4), and \( \mathcal{U}_{x_i^\phi} \), we obtain, at convergence of \( (\hat{\alpha}, \hat{\beta}, \hat{\phi}) \),
\[
\sum_{i=1}^{n} \left( \sum_{i=1}^{n} \left( \frac{\partial \xi^t_i}{\partial \phi} e_i^* \right) \right) \left( \begin{array}{c}
\hat{\alpha} - \alpha \\
\hat{\beta} - \beta \\
\hat{\phi} - \phi
\end{array} \right) - \frac{1}{n} \sum_{i=1}^{n} \left( P_i^* + P_i^{**} \right) e_i^* - \frac{1}{n} \sum_{i=1}^{n} \left( 0 \right) e_i^*
\]

where \( \frac{\partial \hat{\xi}_i}{\partial \phi} = \frac{\partial \xi_i}{\partial \phi} + o(1) \)

by Lemma 3 and \( \Delta_i = \Delta_i + o_p(1) \). Therefore,

\[
\sum_{i=1}^{n} \left( \begin{array}{c}
\hat{\alpha} - \alpha \\
\hat{\beta} - \beta \\
\hat{\phi} - \phi
\end{array} \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left( \begin{array}{c}
P_i \Delta_i \\
-(P_i^* + P_i^{**})
\end{array} \right) e_i^* + \frac{1}{n} \sum_{i=1}^{n} \left( 0 \right) e_i^*
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left( \begin{array}{c}
P_i \Delta_i \\
-(P_i^* + P_i^{**})
\end{array} \right) \tilde{W}_i^{-1} \left( \begin{array}{c}
0 \\
0
\end{array} \right) W_i^{*-1} \left( \begin{array}{c}
y_i - \mu_i \\
y_i^* - \xi_i
\end{array} \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} Q_i \left( \begin{array}{c}
y_i - \mu_i \\
y_i^* - \xi_i
\end{array} \right)
\]
Then by the Central Limit Theorem, we obtain

\[
\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \\ \hat{\phi} - \phi \end{pmatrix} = \mathbf{B}^* - 1 \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{Q}_i \begin{pmatrix} y_i - \mu_i \\ y^*_i - \xi_i \end{pmatrix}
\]

\[
\xrightarrow{d} \mathcal{N}(0, \mathbf{B}^* - 1 \Sigma^* \mathbf{B}^* - 1^T).
\]

Now it remains to show that \( \Sigma^* = \text{cov} \left( \mathbf{Q}_i \begin{pmatrix} y_i - \mu_i \\ y^*_i - \xi_i \end{pmatrix} \right) \), which can be obtained as follows by a direct calculation.

\[
\text{cov} \left( E \left[ \mathbf{Q}_i \begin{pmatrix} y_i - \mu_i \\ y^*_i - \xi_i \end{pmatrix} \bigg| \mathbf{U}_i \right] \right) + E \left( \text{cov} \left[ \mathbf{Q}_i \begin{pmatrix} y_i - \mu_i \\ y^*_i - \xi_i \end{pmatrix} \bigg| \mathbf{U}_i \right] \right)
\]

\[
= 0 + E \left( \mathbf{Q}_i \cdot \text{cov} \left[ \begin{pmatrix} y_i - \mu_i \\ y^*_i - \xi_i \end{pmatrix} \bigg| \mathbf{U}_i \right] \mathbf{Q}_i^T \right) = E(\mathbf{Q}_i \Omega \mathbf{Q}_i^T).
\]

References


Table 1: The Analyses of a Data Set from the Framingham Heart Study

<table>
<thead>
<tr>
<th></th>
<th>Est.</th>
<th>S.E.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1.230</td>
<td>0.251</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age</td>
<td>0.471</td>
<td>0.205</td>
<td>0.022</td>
</tr>
<tr>
<td>HDL</td>
<td>0.471</td>
<td>0.200</td>
<td>0.019</td>
</tr>
<tr>
<td>BMI</td>
<td>0.747</td>
<td>0.184</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Association</td>
<td>0.247</td>
<td>0.192</td>
<td>0.199</td>
</tr>
</tbody>
</table>

Table 2: The biases and standard errors of the proposed estimators for the simulated data.

<table>
<thead>
<tr>
<th>$\psi_{ijk}$</th>
<th>Bias</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$ $\alpha_1$ $\alpha_2$ $\alpha_3$ $\phi$</td>
<td>$\beta$ $\alpha_1$ $\alpha_2$ $\alpha_3$ $\phi$</td>
</tr>
<tr>
<td>0.8</td>
<td>0.089 -0.087 -0.057 -0.072 0.010</td>
<td>0.234 0.313 0.262 0.277 0.110</td>
</tr>
<tr>
<td>1.0</td>
<td>0.088 -0.156 -0.092 -0.049 -0.003</td>
<td>0.235 0.354 0.302 0.314 0.132</td>
</tr>
<tr>
<td>1.5</td>
<td>0.102 -0.100 -0.080 -0.088 0.026</td>
<td>0.225 0.344 0.257 0.335 0.132</td>
</tr>
<tr>
<td>2.0</td>
<td>0.127 -0.062 -0.061 -0.064 0.154</td>
<td>0.202 0.283 0.250 0.262 0.165</td>
</tr>
</tbody>
</table>
Figure 1: Estimated Nonlinear Curves
Table 3: Simulation Results for Model Misspecification: Case 1

<table>
<thead>
<tr>
<th>$\psi_{i,j,k}$</th>
<th>Bias</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$</td>
<td>$\alpha_1$</td>
</tr>
<tr>
<td>0.8</td>
<td>0.095</td>
<td>-0.220</td>
</tr>
<tr>
<td>1.0</td>
<td>0.101</td>
<td>-0.271</td>
</tr>
<tr>
<td>1.5</td>
<td>0.111</td>
<td>-0.268</td>
</tr>
<tr>
<td>2.0</td>
<td>0.109</td>
<td>-0.286</td>
</tr>
</tbody>
</table>

Table 4: Simulations Results for Model Misspecification: Case 2

<table>
<thead>
<tr>
<th>Method</th>
<th>$\psi_{i,j,k}$</th>
<th>Bias</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\beta$</td>
<td>$\alpha_1$</td>
</tr>
<tr>
<td>1</td>
<td>0.8</td>
<td>0.035</td>
<td>-0.029</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.046</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.037</td>
<td>-0.050</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.049</td>
<td>-0.008</td>
</tr>
<tr>
<td>2</td>
<td>0.8</td>
<td>0.033</td>
<td>-0.109</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.033</td>
<td>-0.121</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.037</td>
<td>-0.078</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.032</td>
<td>-0.086</td>
</tr>
</tbody>
</table>