Efficient semiparametric estimation in generalized partially linear additive models for longitudinal/clustered data

GUANG CHENG\textsuperscript{1}, LAN ZHOU\textsuperscript{2,*} and JIANHUA Z. HUANG\textsuperscript{2,**}

\textsuperscript{1}Purdue University, West Lafayette, IN 47907, USA. E-mail: chengg@purdue.edu
\textsuperscript{2}Texas A\&M University, College Station, TX 77843, USA. E-mail: *lzhou@stat.tamu.edu; **jianhua@stat.tamu.edu

We consider efficient estimation of the Euclidean parameters in a generalized partially linear additive models for longitudinal/clustered data when multiple covariates need to be modeled nonparametrically, and propose an estimation procedure based on a spline approximation of the nonparametric part of the model and the generalized estimating equations (GEE). Although the model in consideration is natural and useful in many practical applications, the literature on this model is very limited because of challenges in dealing with dependent data for nonparametric additive models. We show that the proposed estimators are consistent and asymptotically normal even if the covariance structure is misspecified. An explicit consistent estimate of the asymptotic variance is also provided. Moreover, we derive the semiparametric efficiency score and information bound under general moment conditions. By showing that our estimators achieve the semiparametric information bound, we effectively establish their efficiency in a stronger sense than what is typically considered for GEE. The derivation of our asymptotic results relies heavily on the empirical processes tools that we develop for the longitudinal/clustered data. Numerical results are used to illustrate the finite sample performance of the proposed estimators.

Keywords: GEE; link function; longitudinal data; partially linear additive models; polynomial splines

1. Introduction

The partially linear model has become a widely used semiparametric regression model because it provides a nice trade-off between model interpretability and flexibility. In a partially linear model, the mean of the outcome is assumed to depend on some covariates $X$ parametrically and some other covariates $T$ nonparametrically. Usually, the effects of $X$ (e.g., treatment) are of major interest, while the effects of $T$ (e.g., confounders) are nuisance parameters. Efficient estimation for partially linear models has been extensively
studied and well understood for independent data; see, for example, Chen [3], Speckman [21], and Severini and Staniswalis [20]. The book of Härdle, Liang and Gao [8] provides a comprehensive review of the subject.

Efficient estimation of the Euclidean parameter (i.e., the parametric component) in the partially linear model for dependent data is by no means simple due to complication in data structure. Lin and Carroll [16, 17] showed that, whether a natural application of the local polynomial kernel method can yield a semiparametric efficient estimator depends on whether the covariate modeled nonparametrically is a cluster-level covariate or not. Because the naive approach fails, Wang, Carroll and Lin [25] constructed a semiparametric efficient estimator by employing the iterative kernel method of Wang [24] that can effectively account for the within-cluster correlation. Alternatively, Zhang [27], Chen and Jin [4], and Huang, Zhang and Zhou [12] constructed semiparametric efficient estimators by extending the parametric generalized estimating equations (GEE) of Liang and Zeger [15]. He, Zhu and Fung [10] and He, Fung and Zhu [9] considered robust estimation, and Leng, Zhang and Pan [14] studied joint mean-covariance modeling for the partially linear model also by extending the GEE. In all these development, only one covariate is modeled nonparametrically.

In many practical situations, it is desirable to model multiple covariates nonparametrically. However, it is well known that multivariate nonparametric estimation is subject to the curse of dimensionality. A widely used approach for dimensionality reduction is to consider an additive model for the nonparametric part of the regression function in the partly linear model, which in turn results in the partially linear additive model. Although adapting this approach is a natural idea, there are major challenges for estimating the additive model for dependent data. Until only very recently, Carroll et al. [2] gave the first contribution on the partly linear additive model for longitudinal/clustered data, focusing on a simple setup of the problem, where there is the same number of observations per subject/cluster, and the identity link function is used.

The goal of the paper is to give a thorough treatment of the problem in the general setting that allows a monotonic link function and unequal number of observations among subjects/clusters. In this general setting, we derive the semiparametric efficient score and efficiency bound to obtain a benchmark for efficient estimation. In our derivation, we only assume the conditional moment restrictions instead of any distributional assumptions, for example, the multivariate Gaussian error assumption employed in Carroll et al. [2]. It turns out the definition of the efficient score involves solving a system of complex integral equations and there is no closed-form expression. This fact rules out the feasibility of constructing efficient estimators by plugging the estimated efficient influence function into their asymptotic linear expansions. We propose an estimation procedure that approximates the unknown functions by splines and uses the generalized estimating equations. To differentiate our procedure with the parametric GEE, we refer to it as the extended GEE. We show that the extended GEE estimators are semiparametric efficient if the covariance structure is correctly specified and they are still consistent and asymptotically normal even if the covariance structure is misspecified. In addition, by taking advantage of the spline approximation, we are able to give an explicit consistent estimate of the asymptotic variance without solving the system of integral equations that
lead to the efficient scores. Having a closed-form expression for the asymptotic variance is an attractive feature of our method, in particular when there is no closed-form expression of the semiparametric efficiency bound. Another attractive feature of our method is the computational simplicity, there is no need to resort to the computationally more demanding backfitting type algorithm and numerical integration, as has been done in the previous work on the same model.

As a side remark, one highlight of our mathematical rigor is the careful derivation of the smoothness conditions on the least favorable directions from primitive conditions. This rather technical but important issue has not been well treated in the literature. To develop the asymptotic theory in this paper, we rely heavily on some new empirical process tools which we develop by extending existing results from the i.i.d. case to the longitudinal/clustered data.

The rest of the paper is organized as follows. Section 2 introduces the setup of the partially linear additive model and the formulation of the extended GEE estimator. Section 3 lists all regularity conditions, derives the semiparametric efficient score and the efficiency bound, and presents the asymptotic properties of the extended GEE estimators. Section 4 illustrates the finite sample performance of the GEE estimators using a simulation study and a real data. The proofs of some nonasymptotic results and the sketched proofs of the main asymptotic results are given in the Appendix. The supplementary file discusses the properties of the least favorable directions, presents the relevant empirical process tools and the complete proofs of all asymptotic results.

Notation. For positive number sequences $a_n$ and $b_n$, let $a_n \lesssim b_n$ mean that $a_n / b_n$ is bounded, $a_n \asymp b_n$ mean that $a_n \lesssim b_n$ and $a_n \gtrsim b_n$, and $a_n \ll b_n$ mean that $\lim_n a_n / b_n = 0$. For two positive semidefinite matrices $A$ and $B$, let $A \succeq B$ mean that $A - B$ is positive semidefinite. Define $x \lor y$ $(x \land y)$ to be the maximum (minimum) value of $x$ and $y$. For any matrix $V$, denote $\lambda_{V}^{\max}$ $(\lambda_{V}^{\min})$ as the largest (smallest) eigenvalue of $V$. Let $|V|$ denote the Euclidean norm of the vector $V$. Let $\|a\|_{L_2}$ denote the usual $L_2$ norm of a squared integrable function $a$, where the domain of integration and the dominating measure should be clear from the context.

### 2. The model setup

Suppose that the data consist of $n$ clusters with the $i$th ($i = 1, \ldots, n$) cluster having $m_i$ observations. In particular, for longitudinal data a cluster represents an individual subject. The data from different clusters are independent, but correlation may exist within a cluster. Let $Y_{ij}$ and $(X_{ij}, T_{ij})$ be the response variable and covariates for the $j$th ($j = 1, \ldots, m_i$) observation in the $i$th cluster. Here $X_{ij} = (X_{ij1}, \ldots, X_{ijK})'$ is a $K \times 1$ vector and $T_{ij} = (T_{ij1}, \ldots, T_{ijD})'$ is a $D \times 1$ vector. We consider the marginal model

$$
\mu_{ij} = E(Y_{ij}|X_{ij}, T_{ij}),
$$

and the marginal mean $\mu_{ij}$ depends on covariates $X_{ij}$ and $T_{ij}$ through a known monotonic and differentiable link function $\mu(\cdot)$:

$$
\mu_{ij} = \mu(X_{ij}'\beta + \theta_{+}(T_{ij}))
$$
where $\beta$ is a $K \times 1$ vector, and $\theta_+(t)$ is an additive function with $D$ smooth additive component functions $\theta_d(t_d)$, $1 \leq d \leq D$. For identifiability, it is assumed that $\int T_d \theta_d(t_d) \, dt_d = 0$, where $T_d$ is the compact support of the covariate $T_{ijd}$. Applications of marginal models for longitudinal/clustered data are common in the literature (Diggle et al. [7]).

Denote

$$
\begin{align*}
\mathbf{Y}_i &= \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{imi} \end{pmatrix}, & \mathbf{\mu}_i &= \begin{pmatrix} \mu_{i1} \\ \vdots \\ \mu_{imi} \end{pmatrix}, & \mathbf{X}_i &= \begin{pmatrix} \mathbf{X}_{i1}' \\ \vdots \\ \mathbf{X}_{imi}' \end{pmatrix}, & \mathbf{T}_i &= \begin{pmatrix} \mathbf{T}_{i1}' \\ \vdots \\ \mathbf{T}_{imi}' \end{pmatrix}, \\
\theta_+(\mathbf{T}_i) &= \begin{pmatrix} \theta_+(T_{i1}) \\ \vdots \\ \theta_+(T_{imi}) \end{pmatrix}, & \mu(\mathbf{X}_i, \theta_+(\mathbf{T}_i)) &= \begin{pmatrix} \mu(\mathbf{X}_{i1}', \theta_+(\mathbf{T}_{i1})) \\ \vdots \\ \mu(\mathbf{X}_{imi}', \theta_+(\mathbf{T}_{imi})) \end{pmatrix}.
\end{align*}
$$

Here and hereafter, we make the notational convention that application of a multivariate function to a matrix is understood as application to each row of the matrix, and similarly application of a univariate function to a vector is understood as application to each element of the vector. Using matrix notation, our model representation (1) and (2) can be written as

$$
\mathbf{\mu}_i = E(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{T}_i) = \mu(\mathbf{X}_i, \theta_+(\mathbf{T}_i)).
$$

Note that in our modeling framework no distributional assumptions are imposed on the data other than the moment conditions specified in (1) and (2). In particular, $\mathbf{X}_i$ and $\mathbf{T}_i$ are allowed to be dependent, as commonly seen for longitudinal/clustered data.

Let $\Sigma_i = \text{var}(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{T}_i)$ be the true covariance matrix of $\mathbf{Y}_i$. Following the generalized estimating equations (GEE) approach of Liang and Zeger [15], we introduce a working covariance matrix $\mathbf{V}_i = \mathbf{V}_i(\mathbf{X}_i, \mathbf{T}_i)$ of $\mathbf{Y}_i$, which can depend on a nuisance finite-dimensional parameter vector $\tau$ distinct from $\beta$. In the parametric setting, Liang and Zeger [15] showed that, consistency of the GEE estimator is guaranteed even when the covariance matrices are misspecified, and estimation efficiency will be achieved when the working covariance matrices coincide with the true covariance matrices, that is, when $\mathbf{V}_i(\tau^*) = \Sigma_i$ for some $\tau^*$. In this paper, we shall establish a similar result in a semiparametric context.

To estimate the functional parameters, we use basis approximations (e.g., Huang, Wu and Zhou [11]). We approximate each component function $\theta_d(t_d)$ of the additive function $\theta_+(t)$ in (2) by a basis expansion, that is,

$$
\theta_d(t_d) \approx \sum_{q=1}^{Q_d} \gamma_{dq} B_{dq}(t_d) = \mathbf{B}_{dq}'(t_d) \gamma_d,
$$

where $B_{dq}(\cdot), q = 1, \ldots, Q_d$, is a system of basis functions, which is denoted as a vector $\mathbf{B}_d(\cdot) = (B_{d1}(\cdot), \ldots, B_{dQ_d}(\cdot))'$, and $\gamma_d = (\gamma_{d1}, \ldots, \gamma_{dQ_d})'$ is a vector of coefficients. In prin-
Partially linear additive models for longitudinal data

5

ciple, any basis system can be used, but B-splines are used in this paper for their good approximation properties. In fact, if \( \theta_d(\cdot) \) is continuous, the spline approximation can be chosen to satisfy \( \sup_t |\theta_d(t) - B_d(t)\gamma_d| \to 0 \) as \( Q_d \to \infty \), and the rate of convergence can be characterized based on the smoothness of \( \theta_d(\cdot) \); see de Boor [6].

It follows from (4) that

\[
\theta_+ (T_{ij}) \approx \sum_{d=1}^D \sum_{q=1}^{Q_d} \gamma_{dq} B_d(T_{ij}) = \sum_{d=1}^D B_d'(T_{ijd}) \gamma_d = Z_{ij} \gamma,
\]

(5)

where \( Z_{ij} = (B_1'(T_{ij}), \ldots, B_D'(T_{ijd}))' \), and \( \gamma = (\gamma_1', \ldots, \gamma_D')' \). Denoting \( Z_i = (Z_{i1}, \ldots, Z_{im})' \), (3) and (5) together imply that

\[
\mu_i = E(Y_i|X_i, T_i) \approx \mu(X_i \beta + Z_i \gamma).
\]

(6)

Thus, the Euclidean parameters and functional parameters are estimated jointly by minimizing the following weighted least squares criterion

\[
\sum_{i=1}^n (Y_i - \mu(X_i \beta + Z_i \gamma))' V_i^{-1} (Y_i - \mu(X_i \beta + Z_i \gamma))
\]

(7)

or, equivalently, by solving the estimating equations

\[
\sum_{i=1}^n X_i' \Delta_i V_i^{-1} (Y_i - \mu(X_i \beta + Z_i \gamma)) = 0
\]

(8)

and

\[
\sum_{i=1}^n Z_i' \Delta_i V_i^{-1} (Y_i - \mu(X_i \beta + Z_i \gamma)) = 0,
\]

(9)

where \( \Delta_i \) is a diagonal matrix with the diagonal elements being the first derivative of \( \mu(\cdot) \) evaluated at \( X_i \beta + Z_i \gamma \), \( j = 1, \ldots, m_i \). Denoting the minimizer of (7) as \( \hat{\beta} \) and \( \hat{\gamma} \), then \( \hat{\beta} \) estimates the parametric part of the model, and \( \hat{\theta}_1(\cdot) = B_1'(\cdot) \hat{\gamma}_1, \ldots, \hat{\theta}_D(\cdot) = B_D'(\cdot) \hat{\gamma}_D \) estimate the nonparametric part of the model. We refer to these estimators the extended GEE estimators. In this paper, we shall show that, under regularity conditions, \( \hat{\beta} \) is asymptotically normal and, if the correct covariance structure is specified, it is semiparametric efficient, and also show that \( \hat{\theta}_d(\cdot) \) is a consistent estimator of the true nonparametric function \( \theta_d(\cdot) \), \( d = 1, \ldots, D \).

When the link function \( \mu(\cdot) \) is the identity function, the minimizer of the weighted least squares (7) or the solution to the estimating equations (8) and (9) has a closed-from expression:

\[
\begin{bmatrix}
\hat{\beta} \\
\hat{\gamma}
\end{bmatrix}
= \left( \sum_{i=1}^n U_i' V_i^{-1} U_i \right)^{-1} \sum_{i=1}^n U_i' V_i^{-1} Y_i,
\]

where \( U_i = (X_i, Z_i) \).
3. Theoretical studies of extended GEE estimators

3.1. Regularity conditions

We state the regularity conditions needed for the theoretical results in this paper. For the asymptotic analysis, we assume that the number of individuals/clusters goes to infinity while the number of observations per individual/cluster remains bounded.

C1. The random variables $T_{ijd}$ are bounded, uniformly in $i = 1, \ldots, n$, $j = 1, \ldots, m_i$ and $d = 1, \ldots, D$. The joint distribution of any pair of $T_{ijd}$ and $T_{ij'd'}$ has a density $f_{jj'd'd'}(t_{ijd}, t_{ij'd'})$ with respect to the Lebesgue measure. We assume that $f_{jj'd'd'}(\cdot, \cdot)$ is bounded away from 0 and infinity, uniformly in $i = 1, \ldots, n$, $j, j' = 1, \ldots, m_i$, and $d, d' = 1, \ldots, D$.

C2. The first covariate is constant 1, that is, $X_{ij1} \equiv 1$. The random variables $X_{ijk}$ are bounded, uniformly in $i = 1, \ldots, n$, $j = 1, \ldots, m_i$ and $k = 2, \ldots, K$. The eigenvalues of $E\{X_{ij}X_{ij}'|T_{ij}\}$ are bounded away from 0, uniformly in $i = 1, \ldots, n$, $j = 1, \ldots, m_i$.

C3. The eigenvalues of true covariance matrices $\Sigma_i$ are bounded away from 0 and infinity, uniformly in $i = 1, \ldots, n$.

C4. The eigenvalues of the working covariance matrices $V_i$ are bounded away from 0 and infinity, uniformly in $i = 1, \ldots, n$.

Conditions similar to C1–C4 were used and discussed in Huang, Zhang and Zhou [12] when considering partially linear models with the identity link. Condition C1 is also used to ensure identifiability of the additive components, see Lemma 3.1 of Stone [22]. Condition C1 implies that the marginal density $f_{ijd}(\cdot)$ of $T_{ijd}$ is bounded away from 0 on its support, uniformly in $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, and $d = 1, \ldots, D$. The condition on eigenvalues in C2 prevents the multicollinearity of the covariate vector $X_{ij}$ and ensures the identifiability of $\beta$. Since we assume that the cluster size (or the number of observations per subject) is bounded, we expect C3 is generally satisfied. Note that a zero eigenvalue of $\Sigma_i$ indicates that there is a perfect linear relation among the residuals from subject $i$, which is unlikely to happen in reality.

Denote the true values of $\beta$ and $\theta_+(t)$ by $\beta_0$ and $\theta_{0,+}(t)$, respectively.

C5. (i) The link function $\mu$ is strictly monotone and has continuous second derivative; (ii) $\inf_s \mu'(s) = c_1 > 0$; (iii) $\mu'$ and $\mu''$ are locally bounded around $x^T\beta_0 + \theta_{0,+}(t)$; (iv) $\mu(\pm v)$ increases slower than $v^L$ as $v \to \infty$ for some $L > 0$.

Denote $e_{ij} = Y_{ij} - \mu_{ij}$ and $e_i = (e_{i1}, \ldots, e_{im_i})'$.

C6. The errors are uniformly sub-Gaussian, that is,

$$\max_{i=1,\ldots,n} M_0^2 E\{\exp(|e_i|^2/M_0^2) - 1|X_i, T_i\} \leq \sigma_0^2 \quad \forall n, \text{ a.s.} \quad (10)$$

for some fixed positive constants $M_0$ and $\sigma_0$. 
Condition C5 on the link function is satisfied in all practical situations. The sub-
Gaussian condition C6 relaxes the strict multivariate Gaussian error assumption, and is common
used in the literature when applying the empirical process theory.

For \( i = 1, \ldots, n \), let \( \Delta_{i0} \) be a diagonal matrix with the \( j \)th diagonal element being the
first derivative of \( \mu(\cdot) \) evaluated at \( X_{ij}\beta_0 + \theta_{0+i}(T_{ij}) \), \( j = 1, \ldots, m_i \). Let \( X_{ik} \) denote the
\( k \)th column of the matrix \( X_i \). For any additive function \( \varphi_+ (t) = \varphi_1(t_1) + \cdots + \varphi_D(t_D) \),
\( t = (t_1, \ldots, t_D)' \), denote \( \varphi_+(T_i) = (\varphi_+(T_{i1}), \ldots, \varphi_+(T_{im}))' \). Let \( \varphi^{\ast}_{k,+}(\cdot) \) be the additive
function \( \varphi_{k,+}(\cdot) \) that minimizes

\[
\sum_{i=1}^{n} E\{\|X_{ik} - \varphi_{k,+}(T_i)\|^2 \Delta_{i0} V_i^{-1} \Delta_{i0} \{X_{ik} - \varphi_{k,+}(T_i)\} \}. \tag{11}
\]

Denote \( \varphi^*_+(T_i) = (\varphi^*_1(T_i), \ldots, \varphi^*_K(T_i)) \) and define

\[
I_V = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E\{\|X_i - \varphi^*_+(T_i)\|^2 \Delta_{i0} V_i^{-1} \Delta_{i0} \{X_i - \varphi^*_+(T_i)\} \}.
\]

C7. The matrix \( I_V \) is positive definite.

Condition C7 is a positive information requirement that ensures the Euclidean pa-
parameter \( \beta \) can be root-\( n \) consistently estimated. When \( V_i \) is specified to be the true
covariance matrix \( \Sigma_i \) for all \( i \), \( \varphi^*_{k,+}(\cdot) \) reduces to the least favorable direction \( \psi^*_k(\cdot) \)
in the definition of efficient score function and \( I_V \) reduces to the efficient information
matrix \( I_{\varphi_0} \); see Section 3.2.

For \( d = 1, \ldots, D \), let \( G_d = \{B_d'(t)\gamma_d \} \) be a linear space of splines with degree \( r \) de-
defined on the support \( T_d \) of \( T_{ijd} \). Let \( G_+ = G_1 + \cdots + G_D \) be the additive spline space.
We allow the dimension of \( G_d \), \( 1 \leq d \leq D \), and \( G_+ \) to depend on \( n \), but such de-
dpendence is suppressed in our notation to avoid clutter. For each spline space, we require
that the knot sequence satisfies the quasi-uniform condition, that is, max \( j, j' \) \((u_{n,j} + r + 1 -
u_{n,j'} - u_{n,j'})/(u_{n,j} + r + 1 - u_{n,j'}) \) is bounded uniformly in \( n \) for knots \( \{u_{n,j} \} \). Let

\[
\rho_n = \max \left\{ \inf_{g \in G_+} \|g(\cdot) - \theta_{0,+}(\cdot)\|_\infty, \max_{1 \leq k \leq K} \sup_{g \in G_+} \|g(\cdot) - \varphi^*_k(\cdot)\|_\infty \right\}
\]

and \( Q_n = \max\{Q_d = \text{dim}(G_d), 1 \leq d \leq D \} \).

C8. (i) \( \lim_{n \to \infty} Q_n^2 \log^4 n/n = 0 \), (ii) \( \lim_{n \to \infty} n \rho_n^4 = 0 \).

Condition C8(i) characterizes the growth rate of the dimension of the spline spaces relative
to the sample size. Condition C8(ii) describes the requirement on the best rate
of convergence that the functions \( \theta_{0,+}(\cdot) \) and \( \varphi^*_k(\cdot) \)’s can be approximated by functions in
the spline spaces. These requirements can be quantified by smoothness conditions on
\( \theta_{0,+}(\cdot) \) and \( \varphi^*_k(\cdot) \)’s, as follows. For \( \alpha > 0 \), write \( \alpha = \alpha_0 + \alpha_1 \), where \( \alpha_0 \) is an integer
and \( 0 < \alpha_1 \leq 1 \). We say a function is \( \alpha \)-smooth, if its derivative of order \( \alpha_0 \) satisfies a
Hölder condition with exponent \( \alpha_1 \). If all additive components of \( \theta_{0,+}(\cdot) \) and \( \varphi^*_k(\cdot) \)’s
are \( \alpha \)-smooth, and the degree \( r \) of the splines satisfies \( r \geq \alpha - 1 \), then, by a standard
result from approximation theory, \( \rho_n \asymp Q_n^{-\alpha} \) for \( \alpha > 1/2 \) (Schumaker [19]). Condition C8 thus can be replaced by the following condition.

C8'. (i) \( \lim_n Q_n^4 \log^4 n/n = 0 \); (ii) additive components of \( \theta_{0,+} (\cdot) \) and \( \varphi_{k,+}^* (\cdot), \ k = 1, 2, \ldots, K \), are \( \alpha \)-smooth for some \( \alpha > 1/2 \); (iii) \( \lim_n Q_n^{4\alpha}/n = \infty \).

Since \( \varphi_{k,+}^* \) is only implicitly defined, it is important to verify its smoothness requirement from primitive conditions. In the supplementary file (Cheng, Zhou and Huang [5]), that is, Section S.1, we shall show that \( \varphi_{k,+}^* (\cdot) \) solves a system of integral equations and its smoothness is implied by smoothness requirements on the joint density of \( X_i \) and \( T_i \).

### 3.2. Semiparametric efficient score and efficiency bound

For estimating the Euclidean parameter in a semiparametric model, the efficiency bound provides a useful benchmark for the optimal asymptotic behavior (e.g., Bickel et al. [1]). In this subsection, we give the semiparametric efficient score and efficient information matrix when the covariance structure is correctly specified. We do not make the normality assumption on the error distribution in the derivations.

The models studied in this paper have more than one nuisance function so that the efficient score function for \( \beta \), denoted as \( \ell_{\beta}^* \), is obtained by projecting onto a sum-space. In Lemma 1 below, we construct \( \ell_{\beta}^* \) by the two-stage projection approach (Sasieni [18]). Recall that \( e_i = Y_i - \mu(X_i \beta + \theta_+(T_i)) \), where \( \theta_+(t) = \theta_1(t) + \cdots + \theta_D(t_D) \). Write \( f_i(x_i, t_i, Y_i - \mu(X_i \beta + \theta_+(t_i))) \) as the joint density of \( (X_i, T_i, Y_i) \) for the \( i \)th cluster. We assume that \( f_i(\cdot, \cdot, \cdot) \) is smooth, bounded and satisfies \( \lim_{|e_i| \to \infty} f_i(\cdot, \cdot, e_i) = 0 \) for all \( j = 1, \ldots, m_i \).

**Lemma 1.** The efficient score has the expression \( \ell_{\beta}^* = (\ell_{\beta,1}^*, \ldots, \ell_{\beta,K}^*)' \) with

\[
\ell_{\beta,k}^* = \sum_{i=1}^n (X_{ik} - \psi_{k,+}^*(t_i))' \Delta_{i0} \Sigma_i^{-1} [Y_i - \mu(X_i \beta_0 + \theta_{0,+}(T_i))], \tag{12}
\]

where \( \psi_{k,+}^*(t) = \sum_{d=1}^D \psi_{kd}^*(t_d) \) satisfies

\[
\sum_{i=1}^n E[(X_{ik} - \psi_{k,+}^*(T_i))' \Delta_{i0} \Sigma_i^{-1} \Delta_{i0} \psi_d(T_{id})] = 0 \tag{13}
\]

for any \( \psi_d(t_d) \in L_2(T_d), \ d = 1, \ldots, D \).

The form of \( \ell_{\beta,k}^* \) when \( D = 1 \) coincides with that derived in the partially linear models, for example, Lin and Carroll [16], under the strict multivariate Gaussian error assumption. In the supplementary material (Cheng, Zhou and Huang [5]), we shall see that \( \psi_{kd}^*(t_d) \)’s (or, more generally, \( \varphi_{kd}^*(t_d) \)’s) solve a Fredholm integral equation of the second kind (Kress [13]), and do not have a closed-form expression. In the same file, we also show that \( \psi_{kd}^*(t_d) \)’s (or, more generally, \( \varphi_{kd}^*(t_d) \)’s) are well defined and have nice properties.
such as boundedness and smoothness under reasonable assumptions on the joint density of $X_i$ and $T_i$. These properties are crucial for the feasibility to construct semiparametric efficient estimators but are not carefully studied in the literature.

The semiparametric efficient information matrix for $\beta$ is

$$I_{\text{eff}} \equiv \lim_{n \to \infty} \frac{1}{n} E(l_i^\star l_i^\star')$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E[(X_i - \psi_i^\star(T_i))^T \Delta_{ii} \Sigma_i^{-1} \Delta_{ii}(X_i - \psi_i^\star(T_i))],$$

where $\psi_i^\star(T_i) = (\psi_{i1}^\star(T_i), \ldots, \psi_{ik_i}^\star(T_i))$. The efficient information matrix $I_{\text{eff}}$ here is the same as the quantity $I_{V}$ in condition C7 when $V_i = \Sigma_i$. In the above result, different subjects/clusters need not have the same number of observations and thus $(X_i, T_i)$ may not be identically distributed. In the special case that $(X_i, T_i)$ are i.i.d., the efficient information can be simplified to

$$I_{\text{eff}} = E[(X_i - \psi_i^\star(T_i))^T \Delta_{ii} \Sigma_i^{-1} \Delta_{ii}(X_i - \psi_i^\star(T_i))]$$

where the $k$th component of $\psi_i^\star$ satisfies

$$E[(X_{ik} - \psi_{ik_i}^\star(T)) \Delta_{ii} \Sigma_i^{-1} \Delta_{ii}(X_{ik} - \psi_{ik_i}^\star(T))] = 0$$

for any $\psi(d, t_d) \in L_2(T_d), d = 1, \ldots, D$.

The function $\psi_i^\star(T_i)$ involved in the efficient information matrix (14) actually corresponds to the least favorable direction (LFD) along $\theta_{0+}(T_i)$ in the least favorable submodel (LFS). To provide an intuitive interpretation, we assume for simplicity that $f_i(e_i|x_i, t_i) \sim N(0, \Sigma_i)$. Given the above distributional assumption, the parametric submodel (indexed by $\varepsilon$) passing through $(\beta_0, \theta_{0+})$ is constructed as

$$\varepsilon \mapsto -\frac{1}{2} \sum_{i=1}^{n} [y_i - \mu_i(\varepsilon)] \Sigma_i^{-1} [y_i - \mu_i(\varepsilon)],$$

where $\mu_i(\varepsilon) = \mu(x_i(\beta_0 + \varepsilon v) + [\theta_{0+}(t_i) + \varepsilon h_+(t_i)])$, for some vector $v \in \mathbb{R}^K$ and perturbation direction $h_+(\cdot)$ around $\theta_{0+}(\cdot)$. For any fixed $v$, the information matrix for the parametric submodel (evaluated at $\varepsilon = 0$) is calculated as

$$I_{\text{para}}(h_+) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E[(X_i^v + h_+(T_i))^T \Delta_{ii} \Sigma_i^{-1} \Delta_{ii}(X_i^v + h_+(T_i))].$$

The minimum $I_{\text{para}}(h_+)$ over all possible perturbation directions is known as the semiparametric efficient information for $v^T\beta$ (Bickel et al. [1]). The parametric submodel achieving the minimum is called the LFS and the associated direction is called LFD. By calculating the Fréchet derivative of the quadratic function $h_+ \mapsto I_{\text{para}}(h_+)$ and considering (13), we can easily show that its minimum is achieved when $h_+ = -\psi_i^\star v$. In view of the above discussion, the efficient information for $\beta$ becomes the $I_{\text{eff}}$ defined in (14).
Remark 1. Our derivation of the efficient score and efficient information matrix also applies when \( T \) is a cluster level covariate, that is, \( T_{ijd} = T_{id} \) for \( j = 1, \ldots, m_i, \ d = 1, \ldots, D \). Let \( \tilde{T}_i = (T_{i1}, \ldots, T_{iD})' \). In this case, we only need to replace \( \psi_{{k,+}}^*(T_i) \) and \( \psi_{{k,+}}^*(T_i) \) by \( \psi_{{k,+}}^*(T_i)1 \) and \( 1(\psi_{{1,+}}^*(T_i), \ldots, \psi_{{K,+}}^*(T_i)) \), where \( \mathbf{1} \) is an \( m_i \)-vector of ones, and do similar changes for \( \psi_{{k,+}}(T_i) \) and \( \psi_+(T_i) \). It is interesting to note that, when \((Y_i, X_i, T_i)\) are i.i.d., then \( \psi_{{k,+}}^*(\cdot) \) has a closed form expression:

\[
\psi_{{k,+}}^*(t) = \frac{E(X_i'k \Delta_i \Sigma^{-1} \Delta_i 1 | T_i = t)}{E(1' \Delta_i \Sigma^{-1} \Delta_i 1 | T_i = t)}.
\]

3.3. Asymptotic properties

In this subsection, we assume that the dimension of the Euclidean parameter, that is, \( K \), is fixed. Define \( f_0(x, t) = \mu(x' \beta_0 + \theta_{0,+}(t)) \). Define

\[
F_n = \{ f(x, t) : f(x, t) = \mu(x' \beta + g(t), \beta \in \mathbb{R}^K, g \in \mathbb{G}_+ \}.
\]

The extended GEE estimator can be written as

\[
\arg \min_{f \in F_n} \frac{1}{n} \sum_{i=1}^{n} \{ Y_i - f(X_i, T_i) \}' V_i^{-1} \{ Y_i - f(X_i, T_i) \}.
\]

The minimizer is \( \hat{f}_n(x, t) = \mu(x' \hat{\beta}_V + \hat{\theta}(t)) \) where \( \hat{\theta}(t) = \mathbf{B}'(t) \hat{\gamma} \). The subscript of \( \hat{\beta}_V \) denotes the dependence on the working covariance matrices.

According to condition C8 (or C8'), there is an additive spline function \( \theta_n^*(t) = \mathbf{B}'(t) \gamma^* \in \mathbb{G}_+ \) such that \( \| \theta_n^* - \theta_{0,+} \|_{\infty} \leq \rho_n \to 0 \). Then \( f_n(x, t) = \mu(x' \beta_0 + \theta_n^*(t)) \) is a spline-based approximation to the regression function. Define

\[
\langle \xi_1, \xi_2 \rangle_n = \frac{1}{n} \sum_i \xi_1(X_i, T_i) V_i^{-1} \xi_2(X_i, T_i)
\]

and \( \| \xi \|_n^2 = \langle \xi, \xi \rangle_n \).

Theorem 1 (Consistency). The following results hold:

\[
\| \hat{f}_n - f_n^* \|_n^2 = O_P(Q_n \log^2 n / n \vee \rho_n^2), \tag{16}
\]

\[
\| \hat{f}_n - f_n^* \|_{\infty} = o_P(1), \tag{17}
\]

\[
\| \hat{f}_n - f_0 \|_{\infty} = o_P(1), \tag{18}
\]

\[
\hat{\beta}_V \overset{p}{\to} \beta_0, \quad \| \hat{\theta} - \theta_{0,+} \|_{\infty} = o_P(1). \tag{19}
\]

Partially linear additive models for longitudinal data

Theorem 1 says that the extended GEE estimators are consistent in estimating the parametric and nonparametric components of the model. Next we show that, our extended GEE estimator $\hat{\beta}$ is asymptotically normal even when the working covariance matrices $V_i$’s are not necessarily the same as the true ones.

Denote $U_i = (X_i, Z_i)$ as before. Let

$$H = \sum_{i=1}^{n} U_i' \Delta_{i0} V_i^{-1} \Delta_{i0} U_i = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$$

(20)

$$= \begin{pmatrix} \sum_{i=1}^{n} X_i' \Delta_{i0} V_i^{-1} \Delta_{i0} X_i & \sum_{i=1}^{n} X_i' \Delta_{i0} V_i^{-1} \Delta_{i0} Z_i \\ \sum_{i=1}^{n} Z_i' \Delta_{i0} V_i^{-1} \Delta_{i0} X_i & \sum_{i=1}^{n} Z_i' \Delta_{i0} V_i^{-1} \Delta_{i0} Z_i \end{pmatrix}.$$  

By the block matrix form of matrix inverse,

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}^{-1} = \begin{pmatrix} H_{11}^{-1} & H_{12}^{-1} \\ H_{21} & H_{22}^{-1} \end{pmatrix},$$

(21)

where $H_{11,2} = H_{11} - H_{12} H_{22}^{-1} H_{21}$ and $H_{22,1} = H_{22} - H_{21} H_{11}^{-1} H_{12}$. Define

$$R^\Delta(\hat{\beta}_V) \equiv H_{11}^{11} \sum_{i=1}^{n} \{ (X_i - Z_i H_{22}^{-1} H_{21})' \Delta_{i0} V_i^{-1} \Sigma_i V_i^{-1} \Delta_{i0} (X_i - Z_i H_{22}^{-1} H_{21}) \} H_{11}^{11},$$

where the superscript $\Delta$ denotes the dependence on $\Delta_{i0}$.

**Theorem 2 (Asymptotic normality).** The extended GEE estimator $\hat{\beta}_V$ is asymptotically linear, that is,

$$\hat{\beta}_V = \beta_0 + H_{11}^{11} \sum_{i=1}^{n} (X_i - Z_i H_{22}^{-1} H_{21})' \Delta_{i0} V_i^{-1} e_i + o_P \left( \frac{1}{\sqrt{n}} \right).$$

(22)

Consequently,

$$\{R^\Delta(\hat{\beta}_V)\}^{-1/2} (\hat{\beta}_V - \beta_0) \xrightarrow{d} \text{Normal}(0, \mathbf{I}_d),$$

(23)

where $\mathbf{I}_d$ denotes the $K \times K$ identity matrix.

When applying the asymptotic normality result for asymptotic inference, the variance $R^\Delta(\hat{\beta}_V)$ can be estimated by replacing $\Sigma_i$ with $(Y_i - X_i \hat{\beta}_V - Z_i \hat{\gamma})(Y_i - X_i \hat{\beta}_V - Z_i \hat{\gamma})'$, and substituting parameter estimates in $\Delta_{i0}$. The resulting estimator of variance is referred to as the Sandwich estimator.
Theorem 3. $R^\delta(\hat{\beta}_V) \geq R^\delta(\hat{\beta}_\Sigma)$.

Theorem 3 says that $\hat{\beta}_\Sigma$ is the most efficient in the class of extended GEE estimators with general working covariance matrices. Such a result is in parallel to that for standard parametric GEE estimators (Liang and Zeger [15]). This theorem is a consequence of the generalized Cauchy–Schwarz inequality and can be proved using exactly the same argument as Theorem 1 of Huang, Zhang and Zhou [12].

When the covariance matrices are correctly specified, the extended GEE estimators are efficient in a stronger sense than just described. Next, we show that the extended GEE estimator of $\beta$ is the most efficient one among all regular estimators (see Bickel et al. [1] for the precise definition of regular estimators). In other words, the asymptotic variance of $\hat{\beta}_\Sigma$ achieves the semiparametric efficiency bound, that is, the inverse of the efficient information matrix.

Corollary 1. The estimator $\hat{\beta}_\Sigma$ is asymptotically normal and semiparametric efficient, that is,

$$(nI_{eff})^{1/2}(\hat{\beta}_\Sigma - \beta_0) \xrightarrow{d} \text{Normal}(0, I).$$

In the below, we sketch the proof of Corollary 1 and postpone the details to the Appendix. Fixing $V_i = \Sigma_i$ in the definition of $H$ as given in (20), we see that $R^\delta(\hat{\beta}_\Sigma)$ can be written as

$$H^{11}\sum_{i=1}^{n}(\Delta_i \Sigma_i^{-1} \Delta_i)^{\top} (X_i - Z_i H_{22}^{-1} H_{21}) H^{11}.$$  

Using the block matrix inversion formula (21) and examining the (1,1)-block of the identity $H^{-1} = H^{-1} I H^{-1}$, we obtain that $R^\delta(\hat{\beta}_\Sigma) = H^{11}$. Denote $\hat{I}_n = nR^\delta(\hat{\beta}_\Sigma)$. It is easily seen using (21) that

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \Delta_i \Sigma_i^{-1} \Delta_i Z_i$$

$$-\frac{1}{n} \sum_{i=1}^{n} X_i \Delta_i \Sigma_i^{-1} \Delta_i Z_i \left( \sum_{i=1}^{n} Z_i \Delta_i \Sigma_i^{-1} \Delta_i Z_i \right)^{-1} \sum_{i=1}^{n} Z_i \Delta_i \Sigma_i^{-1} \Delta_i X_i.$$  

The asymptotic normality result in Theorem 2 can be rewritten as

$$(n\hat{I}_n)^{1/2}(\hat{\beta}_\Sigma - \beta_0) \xrightarrow{d} \text{Normal}(0, I).$$

Thus, Corollary 1 follows from Theorem 2 and the result that $\hat{I}_n \xrightarrow{p} I_{eff}$. The matrix $\hat{I}_n$ can be interpreted as a spline-based consistent estimate of the efficient information matrix.
Remark 2. When $T$ is a cluster-level covariate, that is, $T_{ij} = \tilde{T}_i = (T_{i1}, \ldots, T_{iD})'$, $j = 1, \ldots, m_i$, Theorems 1, 2 and Corollary 1 still hold. In that case, We can simplify C1 to the following condition.

$C1'$. The random variables $T_{id}$ are bounded, uniformly in $i = 1, \ldots, n$, and $d = 1, \ldots, D$. The joint distribution of any pair of $T_{id}$ and $T_{id'}$ has a density $f_{idd'}(t_{id}, t_{id'})$ with respect to the Lebesgue measure. We assume that $f_{idd'}(\cdot, \cdot)$ is bounded away from 0 and infinity, uniformly in $i = 1, \ldots, n$, and $d, d' = 1, \ldots, D$.

Remark 3. Our asymptotic result on estimation of the Euclidean parameter is quite insensitive to the choice of the number of terms $Q_d$ in the basis expansion which plays the role of a smoothing parameter. Specifically, suppose that additive components of $\theta_{0,\tau}(\cdot)$ and $\varphi_{k,\tau}(\cdot)$, $k = 1, \ldots, K$, all have bounded second derivatives, that is, condition $C8'$ is satisfied with $\alpha = 2$. Then the requirement on $Q_d$ reduces to $n^{1/8} \ll Q_d \ll n^{1/2}/\log^2 n$, a wide range for choosing $Q_d$. Thus, the precise determination of $Q_d$ is not of particular concern when applying our asymptotic results. This insensitivity of smoothing parameter is also confirmed by our simulation study. In practice, it is advisable to use the usual data driven methods such as delete-cluster(subject)-out cross-validation to select $Q_d$ and then check the sensitivity of the results (Huang, Zhang and Zhou [12]).

Remark 4. For simplicity, we assume in our asymptotic analysis that the working correlation parameter vector $\tau$ in $V_i$ is known. It can be estimated via the method of moments using a quadratic function of $Y_i$'s, just as in the application of the standard parametric GEEs (Liang and Zeger [15]). Similar to the parametric case, as long as such an estimate of $\tau$ converges in probability to some $\tau^\dagger$ at $\sqrt{n}$ rate, there is no asymptotic effect on $\hat{\beta}$ due to the estimation of $\tau$; see Huang, Zhang and Zhou [12], Remark 1.

Remark 5. Our method does not require the assumption of normal error distribution. However, because it is essentially a least squares method, it is not robust to outliers. To achieve robustness to outlying observations, it is recommended to use the M-estimator type method as considered in He, Fung and Zhu [9].

4. Numerical results

4.1. Simulation

We conducted simulation studies to evaluate the finite sample performance of the proposed method. When the number of observations are the same per subject/cluster and the identity link function is used, our method performs comparably to the method of Carroll et al. [2] (see supplementary materials). In this section, we focus on simulation setups that cannot be handled by the existing method of Carroll et al. [2]. We generated data from the model

$$E(Y_{ij}|X_{ij}, Z_{ij1}, Z_{ij2}) = g\{\beta_0 + X_{ij}\beta_1 + f_1(Z_{ij1}) + f_2(Z_{ij2})\}, \quad j = 1, \ldots, n, i = 1, \ldots, n,$$
where $g$ is a link function which will be specified below, $\beta_0 = 0$, $\beta_1 = 0.5$, $f_1(t) = \sin\{2\pi(t - 0.5)\}$, and $f_2(t) = t - 0.5 + \sin\{2\pi(t - 0.5)\}$. The covariates $Z_{ij1}$ and $Z_{ij2}$ were generated from independent Normal$(0.5, 0.25)$ random variables but truncated to the unit interval $[0, 1]$. The covariate $X_{ij}$ was generated as $X_{ij} = 3(1 - 2Z_{ij1})(1 - 2Z_{ij2}) + u_{ij}$ where $u_{ij}$ were independently drawn from Normal$(0, 0.25)$. We obtained different simulation setups by varying the observational time distribution, the correlation structure, the parameters of the correlation function, the data distribution, and the number of subjects. We present results for five different setups, the details of which are given below.

For each simulation setup, 400 simulation runs were conducted and summary statistics of the results were calculated. For each simulated data set, the proposed generalized GEE estimator was calculated using a working independence (WI), an exchangeable (EX) correlation, or an autoregressive correlation structure. The correlation parameter $\rho$ was estimated using the method of moments. Cubic splines were used with the number of knots chosen from the range 1–7 by the five-fold delete-subject-out cross-validation. The bias, variance, and the mean squared errors of Euclidean parameters were calculated for each scenario based on the 400 runs. The mean integrated squared errors (MISE), calculated using 100 grip points over $[0, 1]$, for estimating $f_1(\cdot)$ and $f_2(\cdot)$, were also computed.

**Setup 1.** The longitudinal responses are from multivariate normal distribution with the autoregressive correlation structure and the identity link function. For each subject, six observational times are evenly placed between 0 and 1. The results are summarized in Table 1.

**Setup 2.** The same as setup 1, except that the log link is used. The results are summarized in Table 2.

**Setup 3.** This setup is the same as setup 1, except that the exchangeable correlation structure is used and the observational time distribution is different. For each subject, ten observational times are first evenly placed between 0 and 1. Then 40% of the observations are removed from each dataset and thus different subjects may have different number of observations and the observational times may be irregularly placed. The results are summarized in Table 3.

**Setup 4.** It is the same as setup 3, except that the log link is used. The results are summarized in Table 4.

**Setup 5.** This setup is the same as setup 4, except that the Poisson distribution is used as the marginal distribution. All regression parameters in the general setup, the Euclidean and the functional, are halved for appropriate scaling of the response variable. The results are summarized in Table 5.

We have the following observations from the simulation results: for both Euclidean parameters, the estimator accounting for the correlation is more efficient (and sometimes significantly so) than the estimator using working independence correlation structure, even when the correlation structure is misspecified. Using the correct correlation structure usually produces the most efficient estimation. Efficiency gain gets bigger when the correlation parameter $\rho$ gets larger. The variance is usually a dominating factor when comparing the MSEs between the two estimators. We have also observed that the sandwich estimated SEs work reasonably well; the averages of the sandwich estimated SEs
Table 1. Summary of simulation results for setup 1, based on 400 replications. The generalized GEE estimators using a working independence (WI), an exchangeable (EX) correlation structure, and an autoregressive (AR) structure are compared. The true correlation structure is the autoregressive with the lag-one correlation being $\rho$. Each entry of the table equals the original value multiplied by $10^5$.

| $\rho$ | Method | $\beta_0 = 0$ | $\beta_1 = 0.5$ | $f_1(\cdot)$ | $f_2(\cdot)$ |
|--------|---------|---------------|---------------|---------------|---------------|
|        |         | Bias | SD | MSE | Bias | SD | MSE | MISE($f_1$) | MISE($f_2$) |
|        |         | $n = 100$ |      |     |      |     |     |            |             |
| 0.2    | WI      | -27  | 391| 391| -176| 60 | 60 | 1428  | 1330  |
|        | EX      | 14   | 381| 381| -173| 58 | 58 | 1359  | 1307  |
|        | AR      | -27  | 373| 373| -180| 57 | 57 | 1311  | 1279  |
| 0.5    | WI      | -102 | 586| 586| -169| 59 | 60 | 1452  | 1322  |
|        | EX      | 44   | 524| 524| -168| 48 | 49 | 1245  | 1116  |
|        | AR      | -42  | 474| 474| -152| 40 | 40 | 990   | 920   |
| 0.8    | WI      | -194 | 924| 925| -100| 60 | 60 | 1448  | 1358  |
|        | EX      | -13  | 787| 787| -133| 29 | 29 | 747   | 662   |
|        | AR      | -96  | 686| 686| -93 | 16 | 16 | 463   | 461   |
|        | $n = 200$ |      |     |     |      |     |     |            |             |
| 0.2    | WI      | -239 | 181| 182| -35 | 26 | 26 | 689   | 709   |
|        | EX      | -273 | 180| 181| -47 | 25 | 25 | 669   | 698   |
|        | AR      | -238 | 175| 176| -32 | 26 | 26 | 656   | 664   |
| 0.5    | WI      | -261 | 270| 271|  4  | 27 | 27 | 676   | 712   |
|        | EX      | -281 | 258| 259| -38 | 23 | 23 | 569   | 604   |
|        | AR      | -208 | 241| 242| -18 | 19 | 19 | 482   | 493   |
| 0.8    | WI      | -183 | 448| 449|  62 | 30 | 30 | 677   | 723   |
|        | EX      | -243 | 400| 401| -20 | 13 | 13 | 338   | 361   |
|        | AR      | -162 | 369| 369|  -7 |  8 |  8 | 224   | 245   |

are close to the Monte Carlo sample standard deviations (numbers not shown to save space). For the functional parameters $f_1(\cdot)$ and $f_2(\cdot)$, the spline estimator accounting for the correlation is more efficient and the most efficient when the working correlation is the same as the true correlation structure. We also examined the Normal Q–Q plots of the Euclidean parameter estimates and observed that the distributions of the estimates are close to normal. These empirical results agree nicely with our theoretical results.

### 4.2. The longitudinal CD4 cell count data

To illustrate our method on a real data set, we considered the longitudinal CD4 cell count data among HIV seroconverters previously analyzed by Zeger and Diggle [26]. This data set contains 2376 observations of CD4+ cell counts on 369 men infected with the HIV.
Table 2. Summary of simulation results for setup 2, based on 400 replications. The generalized GEE estimators using a working independence (WI), an exchangeable (EX) correlation structure, and an autoregressive (AR) structure are compared. The true correlation structure is the autoregressive with the lag-one correlation being $\rho$. Each entry of the table equals the original value multiplied by $10^5$.

| $\rho$ | Method | $\beta_0 = 0$ Bias | $\beta_0 = 0$ SD | $\beta_0 = 0$ MSE | $\beta_1 = 0.5$ Bias | $\beta_1 = 0.5$ SD | $\beta_1 = 0.5$ MSE | $f_1(\cdot)$ MISE | $f_2(\cdot)$ MISE |
|-------|--------|------------------|------------------|------------------|------------------|------------------|------------------|----------------|----------------|
| 0.2   | WI     | -2294            | 970              | 1022             | 44               | 30               | 30               | 1407           | 2280           |
|       | EX     | -2223            | 962              | 1011             | 77               | 29               | 29               | 1397           | 2195           |
|       | AR     | -2265            | 951              | 1003             | 64               | 29               | 29               | 1374           | 2104           |
| 0.5   | WI     | -2164            | 1137             | 1183             | 62               | 33               | 33               | 1356           | 2198           |
|       | EX     | -1928            | 1007             | 1045             | 117              | 26               | 26               | 1146           | 1846           |
|       | AR     | -1711            | 799              | 828              | 89               | 23               | 23               | 948            | 1394           |
| 0.8   | WI     | -2361            | 1727             | 1783             | 85               | 37               | 37               | 1378           | 2234           |
|       | EX     | -2091            | 1017             | 1061             | 140              | 17               | 17               | 742            | 1303           |
|       | AR     | -1911            | 722              | 758              | 116              | 12               | 12               | 518            | 824            |

|       |        | n = 100          |                  |                  |                  |                  |                  |                |                |
|-------|--------|------------------|------------------|------------------|------------------|------------------|------------------|----------------|----------------|
| 0.2   | WI     | -1387            | 388              | 407              | 88               | 16               | 16               | 601            | 1010           |
|       | EX     | -1498            | 390              | 412              | 99               | 16               | 16               | 582            | 1024           |
|       | AR     | -1497            | 384              | 407              | 98               | 15               | 15               | 565            | 985            |
| 0.5   | WI     | -1433            | 499              | 519              | 84               | 17               | 17               | 618            | 1068           |
|       | EX     | -1532            | 435              | 458              | 108              | 14               | 14               | 534            | 830            |
|       | AR     | -1525            | 387              | 410              | 97               | 12               | 12               | 436            | 660            |
| 0.8   | WI     | -1410            | 712              | 732              | 76               | 18               | 18               | 623            | 1095           |
|       | EX     | -1192            | 332              | 346              | 81               | 9                | 9                | 302            | 482            |
|       | AR     | -1433            | 277              | 298              | 88               | 5                | 5                | 219            | 323            |

virus. See Zeger and Diggle [26] for more detailed description of the data. We fit a partially linear additive model using the log link with the CD4 counts as the response, covariates entering the model linearly including smoking status measured by packs of cigarettes, drug use (yes, 1; no 0), number of sex partners, and depression status measures by the CESD scale (large values indicating more depression symptoms), and the effects of age and time since seroconversion being modeled nonparametrically. We would like to remark that the partially linear additive model here provides a good balance of model interpretability and flexibility. Age and time are of continuous type and thus their effects are naturally modeled nonparametrically. Other variables are of discrete type and are not suitable for a nonparametric model.

Table 6 gives the estimates of the Euclidean parameters using both the WI and EX correlation structures. Cubic splines were used for fitting the additive functions and re-
Table 3. Summary of simulation results for setup 3, based on 400 replications. The generalized GEE estimators using a working independence (WI) and an exchangeable (EX) correlation structure are compared. The true correlation structure is the exchangeable with parameter $\rho$. Each entry of the table equals the original value multiplied by $10^5$.

| $\rho$ | Method | $\beta_0 = 0$ | $\beta_1 = 0.5$ | $f_1(\cdot)$ | $f_2(\cdot)$ |
|-------|--------|---------------|----------------|--------------|--------------|
|       |        | Bias | SD | MSE | Bias | SD | MSE | MISE $(f_1)$ | MISE $(f_2)$ |
| 0     | WI     | $-129$ | $337$ | $337$ | $-125$ | $61$ | $61$ | $1426$ | $1412$ |
|       | EX     | $-109$ | $336$ | $336$ | $-116$ | $61$ | $61$ | $1416$ | $1410$ |
| 0.2   | WI     | $-96$  | $527$ | $527$ | $-56$  | $61$ | $61$ | $1445$ | $1423$ |
|       | EX     | $-161$ | $511$ | $511$ | $-66$  | $55$ | $55$ | $1297$ | $1347$ |
| 0.5   | WI     | $-125$ | $797$ | $797$ | $14$   | $62$ | $62$ | $1515$ | $1399$ |
|       | EX     | $-216$ | $735$ | $735$ | $-39$  | $37$ | $37$ | $924$  | $962$  |
| 0.8   | WI     | $-23$  | $1054$ | $1054$ | $58$   | $62$ | $62$ | $1552$ | $1362$ |
|       | EX     | $-164$ | $914$ | $914$ | $-27$  | $15$ | $15$ | $455$  | $464$  |
|       |        | $n = 200$ |               |               |               |               |               |               |               |
| 0     | WI     | $48$   | $149$ | $149$ | $70$   | $29$ | $29$ | $780$  | $649$  |
|       | EX     | $54$   | $149$ | $149$ | $74$   | $29$ | $29$ | $782$  | $659$  |
| 0.2   | WI     | $-99$  | $253$ | $253$ | $39$   | $29$ | $29$ | $798$  | $677$  |
|       | EX     | $-21$  | $237$ | $237$ | $37$   | $25$ | $25$ | $693$  | $609$  |
| 0.5   | WI     | $-192$ | $403$ | $404$ | $-3$   | $31$ | $31$ | $768$  | $690$  |
|       | EX     | $-64$  | $354$ | $354$ | $15$   | $16$ | $16$ | $470$  | $432$  |
| 0.8   | WI     | $-240$ | $564$ | $565$ | $-60$  | $32$ | $32$ | $718$  | $702$  |
|       | EX     | $-96$  | $466$ | $466$ | $6$    | $7$  | $7$  | $236$  | $227$  |

reported results correspond to the number of knots selected by the five-fold delete-subject-out cross-validation from the range of 0–10. The selected numbers of knots are 8 for time and 4 for age when using the WI structure and 8 for time and 3 for age when using the EX structure. The estimates of the Euclidean parameters using the EX structure have smaller SE than those using the WI structure, suggesting that the EX structure produces more efficient estimates for this data set.

Appendix

A.1. Proof of Lemma 1 (derivation of the efficient score)

Let $\hat{\ell}_\beta$ denote the ordinary score for $\beta$ when only $\beta$ is unknown. Let $\mathcal{P}_f$ and $\mathcal{P}_\theta$ be the models with only $\{f_i, i = 1, \ldots, n\}$ and $\theta_+ (\cdot)$ unknown, respectively, and let $\hat{\mathcal{P}}_f$ and $\hat{\mathcal{P}}_\theta$ be the corresponding tangent spaces. Following the discussions in Section 3.4 of Bickel
Table 4. Summary of simulation results for setup 4, based on 400 replications. The generalized GEE estimators using a working independence (WI) and an exchangeable (EX) correlation structure are compared. The true correlation structure is the exchangeable with parameter $\rho$. Each entry of the table equals the original value multiplied by $10^5$.

| $\rho$ | Method | $\beta_0 = 0$ | $\beta_1 = 0.5$ | $f_1(\cdot)$ | $f_2(\cdot)$ |
|--------|--------|---------------|---------------|---------------|---------------|
|        | Bias   | SD            | MSE           | MISE($f_1$)   | MISE($f_2$)   |
| n = 100|        |               |               |               |               |
| 0      | WI     | $-2451$       | $756$         | $816$         | $-218$        |
|        | EX     | $-2500$       | $773$         | $835$         | $-229$        |
| 0.2    | WI     | $-2581$       | $1054$        | $1120$        | $-176$        |
|        | EX     | $-2440$       | $974$         | $1034$        | $-164$        |
| 0.5    | WI     | $-2555$       | $1514$        | $1574$        | $-88$         |
|        | EX     | $-1970$       | $918$         | $956$         | $-102$        |
| 0.8    | WI     | $-2520$       | $2029$        | $2093$        | $3$           |
|        | EX     | $-2547$       | $1036$        | $1101$        | $24$          |
| n = 200|        |               |               |               |               |
| 0      | WI     | $-883$        | $329$         | $336$         | $114$         |
|        | EX     | $-866$        | $329$         | $337$         | $116$         |
| 0.2    | WI     | $-1090$       | $475$         | $487$         | $84$          |
|        | EX     | $-903$        | $390$         | $398$         | $84$          |
| 0.5    | WI     | $-1310$       | $718$         | $736$         | $44$          |
|        | EX     | $-951$        | $344$         | $353$         | $67$          |
| 0.8    | WI     | $-1533$       | $966$         | $989$         | $13$          |
|        | EX     | $-1245$       | $285$         | $301$         | $65$          |

$\ell_\beta^* = \Pi[\ell_\beta|\hat{P}_f] - \Pi[\Pi(\ell_\beta|\hat{P}_f)\Pi|\hat{P}_f|\hat{P}_f]],$ \hspace{1cm} (25)

where $\Pi[\cdot|\cdot]$ denote the projection operator, and $\hat{P}_f^\perp$ denote the orthogonal complement of $\hat{P}$. Lemma A.4 in Huang, Zhang and Zhou [12] directly implies that

$$\Pi[\ell_\beta|\hat{P}_f^\perp] = \sum_{i=1}^{n} X_i\Sigma^{-1}_{i}\{Y_i - \mu(X_i\beta_0 + \theta_0,+(T_i))\}. \hspace{1cm} (26)$$

Similarly, by constructing parametric submodels for each $\theta_k(\cdot)$ and slightly adapting the same Lemma, we have

$$\Pi[\hat{P}_\theta|\hat{P}_f] = \sum_{i=1}^{n} \left( \sum_{d=1}^{D} \psi_d(T_{id}) \right) \Delta_{i0}\Sigma^{-1}_{i}\{Y_i - \mu(X_i\beta_0 + \theta_0,+(T_i))\}. \hspace{1cm} (27)$$
Table 5. Summary of simulation results for setup 5, based on 400 replications. The generalized GEE estimators using a working independence (WI) and an exchangeable (EX) correlation structure are compared. The true correlation structure is the exchangeable with parameter $\rho$. Each entry of the table equals the original value multiplied by $10^5$.

| $\rho$ | Method | $\beta_0 = 0$ | $\beta_1 = 0.5$ | $f_1(\cdot)$ | $f_2(\cdot)$ |
|--------|--------|---------------|----------------|--------------|--------------|
|        | Bias   | SD      | MSE     | Bias  | SD      | MSE     | MISE($f_1$) | MISE($f_2$) |
| 0      | WI     | $-2967$ | 366     | $-379$ | 113     | 114     | 1376     | 1446      |
|        | EX     | $-2983$ | 368     | $-353$ | 112     | 113     | 1368     | 1439      |
| 0.2    | WI     | $-2557$ | 738     | $-456$ | 120     | 122     | 1394     | 1385      |
|        | EX     | $-2998$ | 777     | $-367$ | 98      | 99      | 1031     | 1110      |
| 0.5    | WI     | $-1952$ | 1101    | 1140   | 221     | 126     | 126      | 1446      |
|        | EX     | $-2272$ | 1339    | 1390   | 215     | 70      | 70       | 506       |
| 0.8    | WI     | $-1979$ | 1344    | 1383   | 369     | 126     | 127      | 1464      |
|        | EX     | $-2349$ | 1651    | 1706   | 506     | 71      | 74       | 411       |

Table 6. Estimates of the Euclidean parameters in the CD4 cell counts study using the spline-based estimates. Working correlation structures used are working independence (WI) and exchangeable (EX). The standard errors (SE) are calculated using the sandwich formula.

| Parameter       | WI Estimate | WI SE   | EX Estimate | EX SE   |
|-----------------|-------------|---------|-------------|---------|
| Smoking         | 0.0786      | 0.0119  | 0.0619      | 0.0111  |
| Drug            | 0.0485      | 0.0421  | 0.0134      | 0.0294  |
| Sex partners    | $-0.0056$   | 0.0043  | $0.0017$    | 0.0035  |
| Depression      | $-0.0025$   | 0.0014  | $-0.0031$   | 0.0013  |
where \( \psi_d(T_{id}) = (\psi_d(T_{i1d}), \ldots, \psi_{im_d}(T_{imi,d}))' \), for \( \psi_d(\cdot) \in L_2(T_d) \). Combination of (25)–(27) gives (12).

### A.2. Proof sketch for Theorem 1 (consistency)

Let \( \epsilon_n = (Q_n/n)^{1/2} \log n \vee \rho_n \). To show (16), it suffices to show that \( P(\| \hat{f}_n - f^n_0 \|_n > \epsilon_n) \to 0 \) as \( n \to \infty \). Applying the peeling device (see the proof of Theorem 9.1 of van de Geer [23]), we can bound the above probability by the sum of \( 2C_0 \exp(-n\epsilon_n^2/(256C_0^2)) \) and \( P(\| y - f^n_0 \|_n > \sigma) \) for some positive constant \( C_0 \). Considering condition C8 and choosing some proper \( \sigma \) related to \( \rho_n \), we complete the proof of (16). As for (17), we have that

\[
\| \hat{f}_n - f^n_0 \|_\infty \leq \| x' \hat{\beta}_V + \hat{\theta} - (x' \beta_0 + \theta^n_0) \|_\infty \leq Q_n^{1/2} \| x' \hat{\beta}_V + \hat{\theta} - (x' \beta_0 + \theta^n_0) \|
\]

by Condition C5(iii) and Lemma S.2 in the supplementary note that

\[
\| x' \beta + g(t) \|_\infty \leq Q_n^{1/2} \| x' \beta + g(t) \| \quad \text{for } g \in \mathbb{G}_+.
\] (28)

It then follows by condition C5(ii) and (16) that

\[
Q_n^{1/2} \| x' \hat{\beta}_V + \hat{\theta} - (x' \beta_0 + \theta^n_0) \| \leq Q_n^{1/2} O_P \{ (Q_n/n)^{1/2} \log n + \rho_n \} = o_P(1)
\]

since \( (Q_n \log n)^2/n \to 0 \) and \( Q_n \rho_n^2 \to 0 \) by condition C8 and the fact that \( \rho_n \asymp Q_n^{-\alpha} \) for \( \alpha > 1/2 \). We thus obtain (17). Due to condition C5(iii), it follows that \( \| f_n^* - f_0 \|_\infty = O(\| \theta^n_0 - \theta_{0,+} \|_\infty) = O(\rho_n) \) by Taylor’s theorem. Combining this with (17), we obtain (18).

From the proof of (17), we have that

\[
\| x' \hat{\beta}_V + \hat{\theta} - (x' \beta_0 + \theta^n_0) \| = O_P \{ (Q_n/n)^{1/2} \log n + \rho_n \}.
\]

Considering Lemma 3.1 of Stone [22], we obtain that \( \| x'(\hat{\beta}_V - \beta_0) \|^2 = o_P(1) \), which together with the no-multicollinearity condition C2 implies \( \hat{\beta}_V \overset{P}{\to} \beta_0 \). By (28), we also obtain

\[
\| \hat{\theta} - \theta^n_0 \|_\infty = Q_n^{1/2} O_P \{ (Q_n/n)^{1/2} \log n + \rho_n \} = o_P(1).
\]

Since \( \| \theta^n_0 - \theta_{0,+} \|_\infty = O(\rho_n) = o(1) \), application of the triangle inequality yields \( \| \hat{\theta} - \theta_{0,+} \|_\infty = o_P(1) \), the last conclusion.

### A.3. Proof sketch for Theorem 2 (asymptotic normality)

Note that \( \hat{\beta}_V \in \mathbb{R}^K \) and \( \hat{\gamma} \in \mathbb{R}^{Q_n} \) solve the estimating equations

\[
\sum_{i=1}^n U_i' \hat{\Delta}_i V_i^{-1} \{ Y_i - \mu(X_i \hat{\beta}_V + Z_i \hat{\gamma}) \} = 0
\] (29)
Partially linear additive models for longitudinal data

with \( U_i = (X_i, Z_i) \), and \( \hat{\Delta}_i \) is a diagonal matrix with the diagonal elements being the first derivative of \( \mu(\cdot) \) evaluated at \( X_i \hat{\beta}_V + Z_i \hat{\gamma} \), \( j = 1, \ldots, m_i \). Using the Taylor expansion, we have that

\[
\mu(X_i \hat{\beta}_V + Z_i \hat{\gamma}) \approx \mu(X_i \beta_0 + Z_i \gamma_0) + \sum_{i=1}^{m_i} \Delta_{i0} \{ X_i (\hat{\beta}_V - \beta_0) + Z_i (\hat{\gamma} - \gamma_0) \}.
\]

Recall that \( \gamma^* \) is assumed to satisfy \( \rho_n = ||\theta_0 + B^\gamma||_\infty \to 0 \). Substituting (30) into (29) yields

\[
0 = \sum_{i=1}^{m_i} \left( \Delta_{i0} \right)^{-1} \left\{ Y_i - \mu(X_i \hat{\beta}_V + Z_i \hat{\gamma}) \right\}.
\]

where

\[
\tilde{J}_1 = (\hat{\Delta}_i - \Delta_{i0}) \left( X_i - Z_i H_{22}^{-1} H_{21} \right) (\hat{\beta}_V - \beta_0)
\]

and

\[
\tilde{J}_2 = \Delta_{i0} \left( Y_i - \mu(X_i \beta_0 + Z_i \gamma_0) - \Delta_{i0} (Z_i \gamma^* - \theta_0) \right).
\]

Recalling (20) and using (21), we obtain from (31) that

\[
\hat{\beta}_V = \beta_0 + H^{11} \sum_{i=1}^{m_i} \left( X_i - Z_i H_{22}^{-1} H_{21} \right) (\tilde{J}_1 + \tilde{J}_2)
\]

\[
= \beta_0 + H^{11} \sum_{i=1}^{m_i} \left( X_i - Z_i H_{22}^{-1} H_{21} \right) \Delta_{i0} \left( \hat{\gamma} - \gamma^* \right) + \pi_n,
\]

where the error term \( \pi_n \) has an explicit form and can be shown to be \( o_P(n^{-1/2}) \) (the proof of this part relies heavily on the empirical process theory and is very lengthy). By the asymptotic linear expansion (22), we have

\[
\{ R^\Delta (\hat{\beta}_V) \}^{-1/2} (\hat{\beta}_V - \beta_0)
\]

\[
= \{ R^\Delta (\hat{\beta}_V) \}^{-1/2} \left( H^{11} \sum_{i=1}^{m_i} \left( X_i - Z_i H_{22}^{-1} H_{21} \right) \Delta_{i0} \left( \hat{\gamma} - \gamma^* \right) \right) + o_P(1).
\]

Then by applying the central limit theorem to the above equation and using the fact that

\[
\text{var} \left( \left. H^{11} \sum_{i=1}^{m_i} \left( X_i - Z_i H_{22}^{-1} H_{21} \right) \Delta_{i0} \left( \hat{\gamma} - \gamma^* \right) \right| \{ X_i, T_i \}_{i=1}^{m_i} \right) = R^\Delta (\hat{\beta}_V),
\]

we complete the whole proof of (23).
A.4. Proof of Corollary 1

We only need to show that $\hat{I}_n \rightarrow I_{ef}$. Fix $V_i = \Sigma_i$ in the definitions of $\langle \xi_1, \xi_2 \rangle^\Delta_n$ and $\langle \xi_1, \xi_2 \rangle^\Delta$. Let $\hat{\psi}_{k,n} = \arg \min_{\psi \in G_i} \|x_k - \psi\|^\Delta_n$. From (21), we see that $\hat{I}_n = (H_{11} - H_{12}H_{21}^{-1}H_{21})/n$. Thus, the $(k,k')$th element of $\hat{I}_n$ is $(x_k - \hat{\psi}_{k,n} - \hat{\psi}_{k',n})^\Delta$. On the other hand, by (13) and (14), the $(k,k')$th element of $I_{ef}$ is the limit of $(x_k - \hat{\psi}_{k,n}^* - x_{k'} - \psi_{k'}^*)_n$, where $\hat{\psi}_{k,n}^* = \psi_{k,n}^* = \arg \min_{L_2, +} \|x_k - \psi\|^\Delta$. Hence, it suffices to show that

$$\|\hat{\psi}_{k,n} - \psi_{k,n}^*\|^\Delta_n = o_P(1), \quad k = 1, 2, \ldots, K, \tag{32}$$

because, if this is true, then by the triangle inequality,

$$\hat{I}_n(k,k') = (x_k - \hat{\psi}_{k,n} - x_{k'} - \hat{\psi}_{k',n})^\Delta_n = (x_k - \hat{\psi}_{k,n}^* - x_{k'} - \psi_{k'}^*)_n + o_P(1) = I_{ef}(k,k') + o_P(1).$$

To show (32), we use $\psi^*_{k,n} = \Pi_{n}^\Delta x_k$ as a bridge. Notice that

$$\|\hat{\psi}_{k,n} - \psi_{k,n}^*\|^\Delta_n \leq \|\psi_{k,n} - \psi_{k,n}^*\|^\Delta_n + \|\hat{\psi}_{k,n} - \psi_{k,n}^*\|^\Delta_n.$$ 

We inspect separately the sizes of the two terms on the right-hand side of the above inequality. First note that $\psi_{k,n}^* = \Pi_{n}^\Delta x_k$ since $G_i \subset L_2, +$. Thus, $\|\psi_{k,n}^* - \psi_{k,n}^*\|^\Delta = \inf_{g \in G_i} \|g - \psi_{k,n}^*\|^\Delta = \inf_{g \in G_i} \|g - \psi_{k,n}^*\|_{L_2} = O(\rho_n) = o(1)$, using Lemma S.2 in the supplementary note. Since $E(\|\psi_{k,n}^* - \psi_{k,n}^*\|^2) = \|\psi_{k,n}^* - \psi_{k,n}^*\|^2$, we have that $\|\psi_{k,n}^* - \psi_{k,n}^*\|^\Delta_n = o_P(1)$. On the other hand, since $\psi_{k,n}^* = \Pi_{n}^\Delta x_k$ and $\hat{\psi}_{k,n} = \Pi_{n}^\Delta x_k$, we have

$$\|\hat{\psi}_{k,n} - \psi_{k,n}^*\|^\Delta_n = \|\psi_{k,n} - \psi_{k,n}^*\|^\Delta_n, \quad \text{and} \quad \|\hat{\psi}_{k,n} - \psi_{k,n}^*\|^\Delta_n \leq \|\psi_{k,n} - \psi_{k,n}^*\|^\Delta_n.$$ 

These two relations and Lemma S.3 in the supplementary note imply that $\|\hat{\psi}_{k,n} - \psi_{k,n}^*\|^\Delta_n = o_P(1)$, which in turn by the same lemma implies $\|\hat{\psi}_{k,n} - \psi_{k,n}^*\|^\Delta_n = o_P(1)$. As a consequence, $\|\hat{\psi}_{k,n} - \psi_{k,n}^*\|^\Delta_n = o_P(1)$, which is exactly (32). The proof is complete.

Acknowledgements

G. Cheng supported by NSF Grant DMS-09-06697 and NSF CAREER Award DMS-1151692. L. Zhou supported in part by NSF Grant DMS-09-07170. J. Z. Huang supported in part by NSF Grants DMS-06-06580, DMS-09-07170, NCI (CA57030), and Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST).

Supplementary Material

Supplement to “Efficient semiparametric estimation in generalized partially linear additive models for longitudinal/clustered data” (DOI: 10.3150/12-BEJ478SUPP). The supplementary file (Cheng, Zhou and Huang [5]) includes the properties of the least favorable directions and the complete
proofs of Theorems 1 and 2 together with some empirical processes results for the clustered/longitudinal data. The results of a simulation study that compares our method with that by Carroll et al. [2] are also included.

References

[1] BICKEL, P.J., KLAASSEN, C.A.J., RITOV, Y. and WELLNER, J.A. (1993). Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins Series in the Mathematical Sciences. Baltimore, MD: Johns Hopkins Univ. Press. MR1245941
[2] CARROLL, R.J., MAITY, A., MAMMEN, E. and YU, K. (2009). Efficient semiparametric marginal estimation for partially linear additive model for longitudinal/clustered data. *Statistics in BioSciences* 1 10–31.
[3] CHEN, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.* 16 136–146. MR0924861
[4] CHEN, K. and JIN, Z. (2006). Partial linear regression models for clustered data. *J. Amer. Statist. Assoc.* 101 195–204. MR2268038
[5] CHENG, G., ZHOU, L. and HUANG, J.Z. (2014). Supplement to “Efficient semiparametric estimation in generalized partially linear additive models for longitudinal/clustered data.” DOI:10.3150/12-BEJ479SUPP.
[6] DE BOOR, C. (2001). *A Practical Guide to Splines*, revised ed. Applied Mathematical Sciences 27. New York: Springer. MR1900298
[7] DIGGLE, P.J., HEAGERTY, P.J., LIANG, K.Y. and ZEGER, S.L. (2002). *Analysis of Longitudinal Data*, 2nd ed. Oxford Statistical Science Series 25. Oxford: Oxford Univ. Press. MR2049007
[8] HÄRDLE, W., LIANG, H. and GAO, J. (2000). *Partially Linear Models*. New York: Springer.
[9] HE, X., FUNG, W.K. and ZHU, Z. (2005). Robust estimation in generalized partial linear models for clustered data. *J. Amer. Statist. Assoc.* 100 1176–1184. MR2236433
[10] HE, X., ZHU, Z.Y. and FUNG, W.K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* 89 579–590. MR1929164
[11] HUANG, J.Z., WU, C.O. and ZHOU, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* 89 111–128. MR1888349
[12] HUANG, J.Z., ZHANG, L. and ZHOU, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. *Scand. J. Stat.* 34 451–477. MR2368793
[13] KRÉSS, R. (1999). *Linear Integral Equations*, 2nd ed. Applied Mathematical Sciences 82. New York: Springer. MR1723850
[14] LENG, C., ZHANG, W. and PAN, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *J. Amer. Statist. Assoc.* 105 181–193. With supplementary material available online. MR2656048
[15] LIANG, K.Y. and ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73 13–22. MR0836430
[16] LIN, X. and CARROLL, R.J. (2001). Semiparametric regression for clustered data. *Biometrika* 88 1179–1185. MR1872228
[17] LIN, X. and CARROLL, R.J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *J. Amer. Statist. Assoc.* 96 1045–1056. MR1947252
[18] Sasieni, P. (1992). Nonorthogonal projections and their application to calculating the information in a partly linear Cox model. *Scand. J. Stat.* **19** 215–233. MR1183198

[19] Schumaker, L.L. (1981). *Spline Functions: Basic Theory.* New York: Wiley. MR0606200

[20] Severini, T.A. and Staniswalis, J.G. (1994). Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.* **89** 501–511. MR1294076

[21] Speckman, P. (1988). Kernel smoothing in partial linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **50** 413–436. MR0970977

[22] Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–171.

[23] van de Geer, S. (2000). *Empirical Processes in M-Estimation.* Cambridge: Cambridge Univ. Press.

[24] Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* **90** 43–52. MR1966549

[25] Wang, N., Carroll, R.J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *J. Amer. Statist. Assoc.* **100** 147–157. MR2156825

[26] Zeger, S.L. and Diggle, P.J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50** 689–699.

[27] Zhang, L. (2004). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. Ph.D. thesis, Univ. Pennsylvania.

*Received October 2011 and revised September 2012*