University of Michigan School of Public Health
The University of Michigan Department of Biostatistics Working Paper Series

Year 2003 *Paper* 11

Efficient Semiparametric Marginal Estimation for Longitudinal/Clustered Data

Naisyin Wang* Raymond J. Carroll†
Xihong Lin‡

*Texas A&M University, nwang@stat.tamu.edu
†Texas A&M University, rcarroll@tamu.edu
‡University of Michigan, xlin@umich.edu

http://biostats.bepress.com/umichbiostat/paper11

Efficient Semiparametric Marginal Estimation for Longitudinal/Clustered Data

Naisyin Wang, Raymond J. Carroll, and Xihong Lin

Abstract

We consider marginal generalized semiparametric partially linear models for clustered data. Lin and Carroll (2001a) derived the semiparametric efficinet score funtion for this problem in the mulitvariate Gaussian case, but they were unable to contruct a semiparametric efficient estimator that actually achieved the semiparametric information bound. We propose such an estimator here and generalize the work to marginal generalized partially liner models. Asymptotic relative efficincies of the estimation or throughout are investigated. The finite sample performance of these estimators is evaluated through simulations and illustrated using a longtiudinal CD4 count data set. Both theoretical and numerical results indicate that properly taking into account the within-subject correlation among the responses can substantially improve efficiency.

Efficient Semiparametric Marginal Estimation for Longitudinal/Clustered Data

By **NAISYIN WANG**, **RAYMOND J. CARROLL** and **XIHONG LIN** [1]

Summary

We consider marginal generalized semiparametric partially linear models for clustered data. Lin and Carroll (2001a) derived the semiparametric efficient score function for this problem in the multivariate Gaussian case, but they were unable to construct a semiparametric efficient estimator that actually achieved the semiparametric information bound. We propose such an estimator here and generalize the work to marginal generalized partially linear models. Asymptotic relative efficiencies of the estimators that ignore the within-cluster correlation structure either in nonparametric curve estimation or throughout are investigated. The finite sample performance of these estimators is evaluated through simulations and illustrated using a longitudinal CD4 count data set. Both theoretical and numerical results indicate that properly taking into account the within-subject correlation among the responses can substantially improve efficiency.

Some key words: Clustered data; Generalized estimating equations; Kernel method; Longitudinal data; Marginal models; Nonparametric regression; Partially linear model; Profile method; Sandwich estimator; Semiparametric information bound; Semiparametric efficient score; Time dependent covariate.

Short title. Efficient Semiparametric Regression for Clustered Data

[1]Naisyin Wang (email: nwang@stat.tamu.edu) is Professor of Statistics and Toxicology and Raymond J. Carroll (email: carroll@stat.tamu.edu) is Distinguished Professor of Statistics, Epidemiology and Biostatistics, Nutrition and Toxicology, Texas A&M University, 3143 TAMU, College Station TX 77843–3143. Xihong Lin (Email: xlin@umich.edu) is Professor of Biostatistics, Department of Biostatistics, School of Public Health, 1420 Washington Heights, Ann Arbor, MI 48109-2029. Wang's research was supported by a grant from the National Cancer Institute (CA74552) and the Texas Advanced Research Program. Carroll's research was supported by a grant from the National Cancer Institute (CA57030) and by the Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30–ES09106). Lin's research was supported by a grant from the National Cancer Institute (CA?????). The authors thank Vincent Carey for sharing his Splus code, "yags", for parametric GEE. We also thank an associate editor and 3 referees for their detail and helpful comments.

1 Introduction

We consider estimation in marginal semiparametric generalized linear models for clustered data using estimating equations. These models are becoming an increasingly popular topic of research, see Zeger and Diggle (1994), Wild and Yee (1996), Pepe and Couper (1997), Hoover, et al. (1998), Lin and Carroll (2001ab) and Lin and Ying (2001) for recent examples.

These marginal models, through general links, have predictor effects that are *partially linear*: they consist of a linear function of one set of predictors (e.g., exposure variables) with a parameter vector β and a completely nonparametric function of a scalar covariate (e.g., time). For uncorrelated data, Severini and Staniswalis (1994) showed how to construct a semiparametric efficient estimator of β using a profile–kernel method. Lin and Carroll (2001a), hereafter referred to as LC, showed that for clustered data, the conventional profile–kernel method does not yield an efficient estimator of β when the parametric covariate is dependent of the nonparametric covariate. In fact, such an estimated β could be $\sqrt{n}$-inconsistent unless either a "*working independence*" (WI) assumption or an under-smoothing step is adopted: here working independence means that one ignores the correlation structure entirely. LC derived the semiparametric efficient score of β in the multivariate Gaussian case, and noted that it was a solution to a complicated Fredholm integral equation. They were however unable to construct an estimator that was semiparametric efficient.

The purpose of this paper is to propose a semiparametric efficient estimator of β in such marginal partially linear models allowing the parametric and nonparametric covariates to be dependent upon one another. When the nonparametric covariate is time, this implies that the parametric covariates could be time dependent. We show that the estimator can effectively account for within-cluster correlation. It is semiparametric efficient in the Gaussian case, and is more efficient than the WI estimator in non–Gaussian cases.

The outline of the paper is as follows. In Section 2, we describe the model and state the

major assumptions. Of particular note is that we are *not* working in the context of time series data: our asymptotics assume that the number of clusters/individuals becomes large while the number of observations per cluster/individual remains bounded. In Section 3 we describe the proposed estimator. Section 4 states the main theoretical results. Numerical studies including an investigation of the asymptotic relative efficiencies of two previously proposed estimators and a small simulation study are provided in Sections 5 and 6, respectively. In Section 7, we analyze a longitudinal data set of CD4 counts of HIV seroconverters. Finally, Section 8 gives concluding remarks.

2 The Model

Suppose that the data consist of n clusters with the ith $(i = 1, \cdots, n)$ cluster having m_i observations. Let Y_{ij} and (X_{ij}, T_{ij}) be the response variable and covariates for the jth $(j = 1, \cdots, m_i)$ observation in the ith cluster. Here X_{ij} is a $p \times 1$ vector and T_{ij} is a scalar that varies within each cluster. Let $\underline{Y}_i = (Y_{i1}, ..., Y_{im_i})^t$ and define $\underline{X}_i$ and $\underline{T}_i$ similarly. Our basic assumption is that the underlying distribution of the response and covariate processes are the same for all subjects, that $(\underline{Y}_i, \underline{X}_i, \underline{T}_i)$ are observations of the ith randomly selected subject within say a fixed range of T such that m_i are bounded, and that

$$E(Y_{ij}|X_{ij}, T_{ij}, \underline{X}_i, \underline{T}_i) = E(Y_{ij}|X_{ij}, T_{ij}) = \mu_{ij}, \tag{1}$$

see Pepe and Couper (1997) for a discussion of this assumption. The marginal mean μ_{ij} depends on X_{ij} and T_{ij} through a known monotonic and differentiable link function $g(\cdot)$:

$$g(\mu_{ij}) = X_{ij}^t \beta + \theta(T_{ij}), \tag{2}$$

where β is a $p \times 1$ vector and $\theta(\cdot)$ is an unknown smooth function. We thus model the effect of X $(p \times 1)$ parametrically and the effect of T nonparametrically. In matrix notation, denoting $\underline{\mu}_i = (\mu_{i1}, \cdots, \mu_{im_i})^t$ and $\underline{g}(\underline{\mu}_i) = \{g(\mu_{i1}), \cdots, g(\mu_{im_i})\}^t$, we have $\underline{g}(\underline{\mu}_i) = \underline{X}_i \beta + \underline{\theta}(\underline{T}_i)$.

As indicated in Section 1, we allow X and T to be dependent. This is in general the case for longitudinal/clustered data. A referee has pointed out the following problem in which the original X and T are independent, but yet can be reparameterized and solved using the proposed method. Specifically, suppose one of the β's, say β_1, in (2) is known to be a linear function of T through $\beta_1(T_{ij}) = \beta_{10} + \beta_{11}T_{ij}$. It is easily seen that $\beta_1(T_{ij})X_{1ij} = \beta_{10}X_{1ij} + \beta_{11}\ X^*_{1ij}$, where $X^*_{1ij} = X_{1ij}T_{ij}$. Thus, model (2) still holds with the added covariate X^*_1, but the X^*_1 is T dependent even if the original X_1 is not. This reparametrization allows us to use the proposed method without modification to obtain inference on β_{10} and β_{11}.

Model (2) differs from a standard marginal GEE model (Liang and Zeger, 1986) mainly by the nonparametric component $\theta(\cdot)$. It is motivated by the fact that the effect of the covariate T (e.g., time) may be complicated and would be better modeled nonparametrically. Applications of marginal models are ample; see Diggle, Liang and Zeger (1994) and Heagerty and Zeger (2000), among others.

Let $\Sigma_i = \Sigma_i(\underline{X}_i, \underline{T}_i)$ and $V_i = V_i(\underline{X}_i, \underline{T}_i)$ be the true and assumed "working" covariances of $\underline{Y}_i$, where $\Sigma_i = \text{var}(\underline{Y}_i|\underline{X}_i, \underline{T}_i)$ and $V_i = S_i^{1/2}R_iS_i^{1/2}$; S_i denotes a diagonal matrix that contains the marginal variances of the Y_{ij}'s, and R_i is an invertible working correlation matrix. Throughout, we assume that V_i can depend on a nuisance finite dimensional parameter vector τ, where τ is distinct from β.

3 The Estimation Procedure

Our estimation procedure is based on profile kernel estimating equations, where $\theta(t)$ is estimated using a kernel GEE estimator accounting for correlations proposed by Wang (2003) and β is estimated using a profile-type estimating equation. The proposed method differs from those proposed by Severini and Staniswalis (1994) and Lin and Carroll (2001a) only in the way that $\widehat{\theta}(t, \beta)$, the estimated $\theta(t)$ for a given β, is constructed. This is motivated

by a fact shown in LC that in order to reach the semiparametric information bound, the within-cluster correlation needs to be properly accounted for in both the parametric and nonparametric estimation procedures. The conventional kernel GEE estimator of $\theta(t)$ (Lin and Carroll, 2000) fails to do so, while new iterative kernel GEE estimator (Wang, 2003) effectively accounts for correlation.

When the link function g is linear, both $\widehat{\beta}$ and $\widehat{\theta}(t,\beta)$ are simply linear estimators. Closed form expressions of the proposed estimators can be shown to exist (Lin, et al, 2003). To better appreciate the nature of the estimators, we present them in an iterative format. For any given β, start with an estimator $\widetilde{\theta}(t,\beta)$ of $\theta(t)$ and an initial estimator $\widetilde{\beta}$ of β satisfying $n^{1/2}(\widetilde{\beta}-\beta) = O_p(1)$. Such initial estimators can be easily obtained, e.g., using the WI estimator that ignores the correlation structure entirely.

We concentrate here on a local linear estimator of $\theta(t)$ proposed by Wang (2003). Let $K_h(s) = h^{-1}K(s/h)$, where K is a mean zero symmetric density function. Define $G_{ij}(t)$ to be an $m_i \times 2$ matrix with the ℓth column $e_j \times \{(t-T_{ij})/h\}^{\ell-1}$ $(\ell = 1,2)$, where e_j is an $m_i \times 1$ vector of 0 except the jth entry being 1. Our method starts with the WI estimator and iterates between steps I and II below until convergence. The working covariance matrix V_i depends on a parameter vector τ, which is assumed to be distinct from β and which can be estimated via the method of moments using quadratic functions of the responses.

- **Step I:** Let $\widetilde{\theta}(\cdot)$ be the current estimator of $\theta(\cdot)$. Given β, let $\widehat{\alpha} = \widehat{\alpha}(t,\beta) = \{\widehat{\alpha}_0(t,\beta), \widehat{\alpha}_1(t,\beta)\}^t$ be the solution to the kernel equation

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(t-T_{ij})\mu_{ij}^{(1)}(\beta,\widehat{\alpha})G_{ij}^t(t)V_i^{-1}[\underline{Y}_i - \mu^*\{t,\underline{X}_i,\underline{T}_i,\beta,\widehat{\alpha},\widetilde{\theta}(\underline{T}_i;\beta)\}] = 0, \qquad (3)$$

where the ℓth element of $\mu^*\{t,\underline{X}_i,\underline{T}_i,\beta,\widehat{\alpha},\widetilde{\theta}(\underline{T}_i,\beta)\}$ is

$$\mu\left[X_{i\ell}^t\beta + I(\ell=j)\{\widehat{\alpha}_0+\widehat{\alpha}_1(t-T_{ij})/h\} + I(\ell\neq j)\widetilde{\theta}(T_{i\ell},\beta)\right],$$

and $\mu_{ij}^{(1)}(\beta, \widehat{\alpha})$ is the first derivative of the function $\mu(\cdot) = g^{-1}(\cdot)$ evaluated at $X_{ij}^t\beta + \widehat{\alpha}_0 + \widehat{\alpha}_1(t - T_{ij})/h$. The updated estimator of $\theta(t)$ is $\widehat{\theta}(t, \beta) = \widehat{\alpha}_0(t, \beta)$.

- **Step II:** Find $\widehat{\beta}$ to solve the following profile-type estimating equation

$$\sum_{i=1}^{n} \frac{\partial \mu\{\underline{X}_i\beta + \widehat{\theta}(\underline{T}_i, \beta)\}^t}{\partial \beta} V_i^{-1} \left[\underline{Y}_i - \mu\{\underline{X}_i\beta + \widehat{\theta}(\underline{T}_i, \beta)\}\right] = 0, \tag{4}$$

Denote by $\{\widehat{\theta}(t), \widehat{\beta}\}$ the estimates at convergence with $\widehat{\theta}(t) = \widehat{\theta}(t, \widehat{\beta})$. As mentioned above, our algorithm differs from those previously proposed in step I by replacing the original kernel GEE estimator by one that utilizes the correlations, while Step II of the algorithm is the same. This modification turns out to be the key to constructing a semiparametric efficient estimator of β.

As shown in Section 4 and later illustrated in the simulation, the proposed $\widehat{\beta}$ is insensitive to the choice of bandwidth. The bandwidth h can be estimated using regular data driven methods such as leaving-one-subject out cross-validation (Silverman and Rice, 1991; Hoover, et al. 1998). Results for higher order local polynomials can be easily obtained by following the derivations given in Ruppert and Wand (1994).

4 Theoretical Results

We emphasize that we assume in our asymptotic theory that the number of clusters $n \to \infty$, while the cluster sizes m_i remains bounded. If m_i also tends to ∞, the problem is quite different, as pointed out by LC. We assume that the regularity conditions in the Appendix hold. Denote by β_0 and $\theta_0(t)$ the true values of β and $\theta(t)$. Let $d^{(r)}(\cdot)$ be the rth derivative of any function $d(\cdot)$, $\alpha_\ell = h^\ell\theta^{(\ell)}(t)/\ell!$, $v_i^{j\ell}$ be the (j, ℓ)th element of a matrix V_i^{-1}, and $f_j(t)$ be the marginal density of the T_{ij}.

Our results concerning $\widehat{\theta}(t)$ are simple: they coincide exactly with those of Wang (2003), who shows that $\widehat{\theta}(t)$ can effectively account for the within-cluster correlation and is asymp-

totically more efficient than the WI estimator. The main focus of this paper is on the properties of $\widehat{\beta}$.

LC gives the semiparametric efficient score for a multivariate normal partially linear model with a known Σ for all i. We use the same setup but allow the conditional mean to be as given in (2) and conditional variance to be $\Sigma_i(\underline{X}_i, \underline{T}_i)$ with known τ. The results remain the same with estimated τ. A discussion of this point is given in Remark 5.

A referees has also suggested that we provide a link of our work to the least favorable direction principle in Bickel, et al. (1993). This link, described in Appendix A.1, provide an alternative derivation of the efficient score than that given by LC, and also allows extension to the unequal m_i case.

Under the assumed multivariate normal structure, we show that $\widehat{\beta}$ is semiparametric efficient (Proposition 1 and Corollary 1). The results are then extended to model (2) for general outcomes with a working covariance V and without distributional assumption (Proposition 2).

Hereafter, we assume that $m_i \equiv m$ and that $(\underline{Y}_i, \underline{X}_i, \underline{T}_i)$ are i.i.d. This allows us to reduce the complexity of notation needed for presentation and concentrate on the main concept. Let $\Delta = \Delta(\underline{X}, \underline{T})$ be a diagonal matrix with the diagonal element being the first derivative of μ. We show in Appendix A.1 that the semiparametric information bound for β under the multivariate normal assumption is

$$E\left[\{\underline{X} - \varphi_{eff}(\underline{T})\}^t \Delta \Sigma^{-1} \Delta \{\underline{X} - \varphi_{eff}(\underline{T})\}\right],$$

where $\varphi_{eff}(\underline{T})$ is an $m \times p$ matrix whose jth row is $\varphi_{eff}(T_j)$, and $\varphi_{eff}(T_j) = \{\varphi_{eff,1}(T_j), \cdots, \varphi_{eff,p}(T_j)\}^T$. The function $\varphi_{eff}(t)$ solves

$$\sum_{j=1}^{m}\sum_{\ell=1}^{m} E\left[\Delta_{jj}\sigma^{j\ell}\Delta_{\ell\ell}\left\{X_\ell - \varphi_{eff}(T_\ell)\right\} | T_j = t\right] f_j(t) = 0, \tag{5}$$

where $\sigma^{j\ell}$ is the (j,ℓ)th element of Σ^{-1} and Δ_{jj} is the(j,j)th element of Δ. Equation (5) corresponds to a Fredholm integral equation of the second kind (Kress, 1989, chap. 1),

namely

$$\varphi_{eff}(t) - \left\{ q(t) - \int H(t,s)\varphi_{eff}(s)ds \right\} = 0, \tag{6}$$

where, denoting $f_{\ell j}(\cdot,\cdot)$ the joint density function of (T_ℓ, T_j), $H(t,s)$ and $q(t)$ are defined as

$$H(t,s) = \frac{\sum_j \sum_{\ell \neq j} E(\Delta_{jj}\sigma^{j\ell}\Delta_{\ell\ell}|T_\ell = s, T_j = t) f_{\ell j}(s,t)}{\sum_{j=1}^m E(\sigma^{jj}\Delta_{jj}^2|T_j = t) f_j(t)}, \tag{7}$$

$$q(t) = \frac{\sum_{j=1}^m \sum_{\ell=1}^m E(\Delta_{jj}\sigma^{j\ell}\Delta_{\ell\ell}X_\ell|T_j = t) f_j(t)}{\sum_{j=1}^m E(\sigma^{jj}\Delta_{jj}^2|T_j = t) f_j(t)}, \tag{8}$$

respectively. Similar calculations were given in LC.

We next study the asymptotic distribution of our estimator $\widehat{\beta}$ and show that it reaches the above semiparametric information bound. Define $\widehat{\varphi}(t,\beta) = -\partial\widehat{\theta}(t,\beta)/\partial\beta^t$. We first show that $\widehat{\varphi}(t,\beta)$ converges to $\varphi_{eff}(t)$, which is a crucial theoretical result of this paper and is the key to the investigation of the asymptotic properties of $\widehat{\beta}$. It is also used later to justify why the proposed estimator does not require under-smoothing of $\theta(t)$ to obtain a $\sqrt{n}$–consistent estimator $\widehat{\beta}$.

Proposition 1 *Let $\widehat{\varphi}$ be the partial derivative of the final estimator of θ w.r.t. β as defined above, and φ be its limit as $n \to \infty$. We have φ satisfies (6), that is, $\varphi(t) = \varphi_{eff}(t)$.*

Corollary 1 *Under the assumed multivariate normal structure, with $h \to 0, n \to \infty$, at the rate that $nh^8 \to 0$, and $nh/log(1/h) \to \infty$, we have*

$$n^{1/2}(\widehat{\beta} - \beta_0) \to Normal\left(0, E\left[\{\underline{X} - \varphi_{eff}(\underline{T})\}^t \Delta\Sigma^{-1}\Delta\{\underline{X} - \varphi_{eff}(\underline{T})\}\right]\right),$$

in distribution, i.e., $\widehat{\beta}$ reaches the semiparametric information bound.

The proofs of Proposition 1 and Corollary 1 are given in Appendices A.3 and A.4, respectively. We now consider the properties of $\widehat{\beta}$ in model (2) for general outcomes assuming a working covariance matrix V and without the normality assumption. The asymptotic properties of $\widehat{\beta}$ are presented in Proposition 2. Recall that $\Delta_i = \Delta(\underline{X}_i, \underline{T}_i) = \text{diag}\{\mu_{ij}^{(1)}\}$, and define $\widetilde{\underline{X}}_i = \underline{X}_i - \varphi(\underline{T}_i, \beta_0)$.

Proposition 2 *Let* $\widetilde{A}(V) = E(\widetilde{\underline{X}}^t \Delta V^{-1} \Delta \widetilde{\underline{X}})$ *and* $\widetilde{B}(V,\Sigma) = E(\widetilde{\underline{X}}^t\ \Delta V^{-1}\ \Sigma V^{-1} \Delta \widetilde{\underline{X}})$. *With* $h \to 0, n \to \infty$, *at the rate that* $nh^8 \to 0$, *and* $nh/log(1/h) \to \infty$, *we have*

$$n^{1/2}(\widehat{\beta} - \beta_0) \to Normal\{0, \Omega(V,\Sigma)\},$$
$$where \qquad \Omega(V,\Sigma) = \left\{\widetilde{A}(V)\right\}^{-1} \widetilde{B}(V,\Sigma) \left\{\widetilde{A}(V)\right\}^{-1}. \tag{9}$$

The asymptotic covariance $\Omega(V,\Sigma)$ *is minimized by* $V = \Sigma$, *and in this case equals* $\widetilde{A}^{-1}(\Sigma)$.

The proof of (9) is sketched in Appendix A.4. Several remarks about our theoretical results are in order.

Remark 1: LC showed that when the conventional profile-kernel method is used, except in the special case where WI is assumed, a $\sqrt{n}$ consistent estimator of β can be obtained only if one artificially under-smooths the nonparametric estimator to eliminate an unwanted bias term. That is, either the bandwidth must be chosen so that $nh^4 \to 0$ or a nonparametric regression algorithm with smaller bias than standard kernel regression must be used, e.g., the twicing method. Even when $\theta(t)$ is undersmoothed, the conventional profile kernel estimator assuming the true correlation is still inefficient. However, using our method, not only is such under-smoothing unnecessary but also the resulting estimator of β is semiparametric efficient.

Remark 2: Under local linear smoothing, Corollary 1 and Proposition 2 hold for a wide range of bandwidths. That is, $\widehat{\beta}$ is quite insensitive to the choice of bandwidth. For example, any data driven methods of order $O_P(n^{-1/5})$ fulfill the requirements. With m_i being finite, Wang (2003) illustrates the following asymptotic phenomenon, namely that the proposed estimate of θ is essentially a locally weighted estimate of cluster-wise pseudo responses, where the ith pseudo response is formed by a linear combination of responses in the ith cluster. This implies that the derivations which justify the use of cross-validation under the independent data scenario can be equivalently carried out here, and that the asymptotic bias and variance

given in Wang can be used to show that the resulting choice of h is of order $n^{-1/5}$, where n is the number of subject.

On the other hand, since $\widehat{\beta}$ is insensitive to the choice of bandwidth, the plug-in bandwidth discussed in Wang (2003), the leaving-one-subject out cross-validation using the WI estimated θ rather than the proposed $\widehat{\theta}$ would all serve the purpose. That is, with the proposed estimating procedure, which data-driven bandwidth to use is not of particular concern, at least asymptotically. An illustration of this phenomenon is provided through a simulation study in Section 6.

Remark 3: Lin and Carroll (2001a) allowed different working covariances in their estimating equations of $\theta(t)$ and β. This is motivated by the fact that the most efficient conventional kernel estimator requires ignoring correlation, while a more efficient estimator of β requires accounting for correlation. A similar approach can be adopted in our method by simply using V_{1i} and V_{2i} to replace V_i in (3) and (4), respectively. The dependence of the result on V_1 is implicitly embedded in $\widetilde{X}$, while $\Omega(V, \Sigma)$ needs to be replaced by $\Omega(V_2, \Sigma)$. However, there is no advantage of doing this in our framework, since our results show that when $V_1 = V_2 = \Sigma$, our method gives the most efficient estimators of both $\theta(t)$ and β. Further, if we allow V_1 and V_2 to be different in our method, a consistent estimator of β requires undersmoothing.

The estimation framework considered by Zeger and Diggle (1994) is a special case of the conventional profile-kernel estimating structure of LC. Specifically, they assumed working independence V_1 for estimating $\theta(t)$ and a non-diagonal V_2 which accounts for the within-cluster correlation for estimating β. LC referred to an extension of their estimator as the "under-smoothed profile-kernel estimator," where $\theta(t)$ is undersmoothed to guarantee the estimator of β to be $\sqrt{n}$-consistent. Even though Zeger and Diggle carry out their calculation with a backfitting method and we really concentrate on kernel profile estimation, hereafter we refer to the estimator of β using working independence V_1 and estimated V_2 as the Zeger-

Diggle (ZD) estimator to credit them for their original idea of choosing the pair (V_1, V_2).

Note that in theory this undersmoothed estimator is still not semiparametric efficient even when assuming $V_2 = \Sigma$. Our numerical asymptotic relative efficiency study in Section 5 suggests that this estimator has a high relative efficiency. However, as pointed out in LC, the consistency of the ZD estimator of β requires under-smoothing in estimating $\theta(t)$. Further, as also pointed out in LC, a practical drawback of this estimator is that a "regular" sandwich variance estimator cannot be utilized. A variance estimator would either involve empirically estimating the complicated Z_1 and Z_2 terms given in Appendix A.5 or a bootstrap method.

Remark 4: Since $\Omega(V, \Sigma)$ is minimized when $V = \Sigma$, i.e., when the correct covariance structure is specified, Proposition 2 implies that the proposed estimator is more efficient than the WI estimator that uses $V = I$, the identity matrix.

Remark 5: Parallel to standard GEEs (Liang and Zeger, 1986), it can be shown that our estimator is still consistent when the working covariance matrix V is misspecified and is most efficient when V is correctly specified. Obviously, more efficiency can be gained by adopting a more complicated estimating equation for β under certain special models such as those with part of τ being β, i.e., there is information for β beyond the mean. This has been done in parametric cases (see, e.g., Prentice and Zhao, 1991; Crowder, 2001). As pointed out in the literature, also shared by our own experience, little information is gained from the added complexity. A relevant discussion is given in Crowder (2001) for parametric cases. In this respect, issues to be considered for proposed estimator are the same to those in parametric models.

For simplicity, we assume in our asymptotic work that the working correlation parameter vector τ in V is known. It can be estimated via the method of moments using a quadratic function of Y's. Following Lin and Carroll (2001b), it can be shown that once such an estimator of τ converges in probability to some τ^* at a $\sqrt{n}$-rate, then there is no asymptotic

effect on our estimators of β and $\theta(\cdot)$ due to estimation of τ, i.e., Proposition 2 still holds. In addition, with τ being of finite dimension, following remark 3.2 of Begun, Hall, Huang and Wellner (1983), it can be shown that the semiparametric information bound given above remains the same whether τ is known or estimated. Consequently, in the case we consider, our estimator of β is semiparametric efficient when $\widehat{\tau}$ is $\sqrt{n}$ consistent to certain τ^*. Following Carroll, Wu and Ruppert (1988), one could iteratively update the estimated τ and no more than 3–4 iterations of this process should suffice even for second–order purposes.

5 Asymptotic Relative Efficiency of Estimated β

In this section, we study numerically the asymptotic relative efficiencies (AREs) of the working independence estimator and the ZD estimator with respect to the semiparametric efficient estimator. We concentrate on the efficiencies of the estimators of β. The ARE of the WI estimator compared to the proposed estimator of $\theta(t)$ is the same as that reported in Wang (2003).

We consider the case where the cluster size is constant ($m_i = m$) and X_{ij} is a scalar Gaussian covariate. The underlying model is $Y_{ij} = X_{ij}\beta + \theta(T_{ij}) + \epsilon_{ij}$. Let $\underline{\epsilon}_i = (\epsilon_{i1}, \cdots, \epsilon_{im})^T$ with $\underline{\epsilon}_i \sim N(0, \Sigma)$, and σ_{jj} be the jth diagonal element of Σ. To simplify calculations, we also assume that T_{ij} is Gaussian even though this violates the assumption that $f_j(t)$ has to be bounded away from 0. One can view the resulting efficiency calculation as an approximation of that when T_{ij} follows a truncated normal. The advantage of assuming normality is that the integral equation (6) has a closed form solution and thus the semiparametric information bound has a closed form. Specifically, we assume both $\underline{X}_i$ $(m \times 1)$ and $\underline{T}_i$ $(m \times 1)$ are centered multivariate normal random variables with mean 0 and covariances $\mathrm{cov}(\underline{X}_i) = \sigma_X^2 \Gamma_{XX}$, $\mathrm{cov}(\underline{T}_i) = \sigma_T^2 \Gamma_{TT}$, $\mathrm{cov}(\underline{X}_i, \underline{T}_i) = \sigma_X \sigma_T \Gamma_{XT}$, where $\Gamma_{XX} = \{\rho_{jk}^{XX}\}$, $\Gamma_{TT} = \{\rho_{jk}^{TT}\}$, and $\Gamma_{XT} = \{\rho_{jk}^{XT}\}$ are the correlation matrices. That is, we assume that all X_{ij} share the

same marginal distribution with a common variance σ_X^2. Similarly, all T_{ij} share the same marginal distribution with a common variance σ_T^2.

We first calculate the semiparametric efficient score by solving the integral equation (6). For simplicity, we suppress the subscript i in the following discussion. Under the above multivariate normal assumption of $\underline{X}$ and $\underline{T}$, calculations sketched in Appendix A.5 show that (6) has a closed form solution

$$\varphi_{eff}(t) = \frac{\sigma_X \sum_{j=1}^m \sum_{k=1}^m \sigma^{jk} \rho_{jk}^{XT}}{\sigma_T \sum_{j=1}^m \sum_{k=1}^m \sigma^{jk} \rho_{jk}^{TT}}\, t, \tag{10}$$

where σ^{jk} is the (j,k)th element of Σ^{-1}. Let $\widetilde{\underline{X}}_{eff} = \underline{X} - \varphi_{eff}(\underline{T})$. The semiparametric information bound of β is $I_{eff}(\beta) = E\left(\widetilde{\underline{X}}_{eff}^T \Sigma^{-1} \widetilde{\underline{X}}_{eff}\right)$.

For the working independence estimator, we assume the working covariance matrix is Σ_d=diag(Σ). Using Result 1 of Lin and Carroll (2001a), the asymptotic information matrices of the working independence (WI) and of the Zeger-Diggle estimator (ZD) are

$$\begin{aligned} I_{WI}(\beta) &= E(\widetilde{\underline{X}}_{WI}^T \Sigma_d^{-1} \widetilde{\underline{X}}_{WI}) E(\widetilde{\underline{X}}_{WI}^T \Sigma_d^{-1} \Sigma \Sigma_d^{-1} \widetilde{\underline{X}}_{WI})^{-1} E(\widetilde{\underline{X}}_{WI}^T \Sigma_d^{-1} \widetilde{\underline{X}}_{WI}), \\ I_{ZD}(\beta) &= E(\widetilde{\underline{X}}_{WI}^T \Sigma^{-1} \widetilde{\underline{X}}_{WI}) E\{(\underline{Z}_1 - \underline{Z}_2)^T \Sigma (\underline{Z}_1 - \underline{Z}_2)\}^{-1} E(\widetilde{\underline{X}}_{WI}^T \Sigma^{-1} \widetilde{\underline{X}}_{WI}), \end{aligned}$$

respectively, where $\widetilde{\underline{X}}_{WI} = \underline{X} - \varphi_{WI}(\underline{T})$, $\varphi_{WI}(t) = \{\sigma_X \sum_{j=1}^m \sigma_{jj}^{-1} \rho_{jj}^{XT}\} / \{\sigma_T \sum_{j=1}^m \sigma_{jj}^{-1}\}\ t$, and $\underline{Z}_1$, $\underline{Z}_2$ are defined in Appendix A.5. The two asymptotic relative efficiencies (AREs) of interest are

$$ARE_{WI}(\beta) = \frac{I_{WI}(\beta)}{I_{eff}(\beta)}, \qquad ARE_{ZD}(\beta) = \frac{I_{ZD}(\beta)}{I_{eff}(\beta)}.$$

It can be easily seen that these asymptotic relative efficiencies are free of the marginal variances of X and T, i.e., of σ_X^2 and σ_T^2.

We performed a numerical asymptotic relative efficiency study by assuming an exchangeable correlation structure on $\underline{Y}$, $\underline{X}$ and $\underline{T}$, i.e., with I being the identity matrix and J being a matrix of 1,

$$\Sigma = \sigma^2\{(1-\rho)I + \rho J\},\ \Gamma_{XX} = (1-\rho_X)I + \rho_X J,\ \Gamma_{TT} = (1-\rho_T)I + \rho_T J.$$

Furthermore, we let $\Gamma_{XT} = \rho_{XT}\{(1-\delta)I + \delta J\}$, with $0 < \delta \leq 1$. That is, $\text{corr}(X_{ij}, T_{ij}) = \rho_{XT}$ and for $j \neq k$, $\text{corr}(X_{ij}, T_{ik}) = \delta\rho_{XT}$, which could be smaller than the correlation between the paired X_{ij} and T_{ij} measured at the same time. Throughout, we set $\delta = 0.6$ and $\rho_T = 0.3$.

Assuming the cluster size $m = 4$, the left and right panels of Figure 1 display the asymptotic relative efficiencies ARE_{WI} and ARE_{ZD} as functions of ρ, the correlation among the outcome $\underline{Y}$. We assume the correlations among $\underline{X}$ and between $\underline{X}$ and $\underline{T}$ to be $\rho_X = \rho_{XT} = 0.3, 0.6$, which represent low and moderate levels of correlation. The results are depicted by the solid (for 0.3) and dotted (for 0.6) curves.

Figure 1 shows that the WI estimator is subject to a moderate amount of efficiency loss even when the correlation among the outcomes Y is modest. The loss of efficiency becomes substantial when the correlation among the outcomes Y becomes large. The ZD estimate which assumes the true correlation in estimating β has a much higher relative efficiency compared to the WI estimator. For example, when $\rho = 0.6$, $\rho_X = \rho_{XT} = 0.3$ and 0.6, ARE_{WI} and ARE_{ZD} are 44.7% and 98.8%, and 47.6% and 92.8%, respectively. Considerable loss of efficiency is found in ZD only when ρ is very large.

The exchange of the relative positions of the two curves in the two panels of Figure 1 suggests that the relationship between the ARE and the level of correlation within and among $\underline{X}$ and $\underline{T}$ differs between the WI and ZD estimators. For example, as the correlation between $\underline{X}$ and $\underline{T}$ increases, the loss of efficiency of the ZD estimator increases, while that of the WI estimator decreases slightly.

Concentrating on the scenario with $\rho_X = \rho_{XT} = 0.6$, we use Figure 2 to illustrate the changes in relative efficiencies with the cluster size m. As in Figure 1, the left and right panels display the asymptotic relative efficiencies ARE_{WI} and ARE_{ZD} as functions of ρ. The curves from the top to the bottom correspond to $m = 3$, 4, 5, and 6, respectively. For both

the WI and ZD estimators, the loss of efficiency increases with the cluster size m.

6 A Simulation Study

In this section, we report a simulation study to investigate the finite sample performance of the proposed estimator and compare it with the WI and ZD estimators. We consider the following longitudinal scenario. For each subject, the time varying covariates $\underline{T}$ and $\underline{X}_1$ were generated as sums of independent uniform $[-1, 1]$ random variables and a common uniform $[0, 1]$ random variable. This made each X_1 and T dependent and gives $\rho_X = \rho_T = \rho_{XT} = 0.2$ and $\delta = 1$. We also included a time independent covariate $\underline{X}_2$, which equals 0 for half of the subjects and 1 for the other half and mimics a treatment indicator. The response Y_{ij} was generated assuming a conditional mean $\mathrm{E}(Y_{ij}|X_{ij}) = \sin(2T_{ij}) + \beta_1 X_{1ij} + \beta_2 X_{2ij}$, a common variance 1 and an exchangeable correlation structure with $\rho = 0.6$. We let $\beta_1 = \beta_2 = 1$, $n = 100$ and $m = 4$. We generated 250 data sets with $\underline{X}_1$ and $\underline{T}$ re-generated each time. An exchangeable correlation structure was assumed with ρ being estimated using the method of moments. All estimates including the profile iterative kernel, WI and ZD methods were computed using the Epanechnikov kernel for K.

To understand how insensitive the proposed estimator is to the choice of the bandwidth h, we did the following. For the first 50 data sets, we estimated the bandwidth using a method mimicking the idea of the empirical bias bandwidth selection method (Ruppert, 1997; Wang, 2003) and the leaving-one-subject-out cross-validation method used in LC. The range of selected bandwidths was then further expanded to $[0.35, 0.65]$. We then evaluated the performance of the three estimators for 7 bandwidths equally spaced between 0.35 and 0.65. For the ZD estimate, each bandwidth was further multiplied by $(n \times m)^{-2/15}$, an under-smoothing required for $\sqrt{n}$–consistency of β estimation. The ratios of the resulting Monte Carlo variances and mean squared errors (MSE's) of each of the three estimates relative to

the proposed efficient estimate as functions of bandwidths are displayed in Figure 3. The top two panels are for the estimates of β_1, while the bottom two are for the estimates of β_2. The solid, dotted and dashed curves correspond to the proposed, WI and ZD estimates, respectively. These results show the estimates of β perform about equally well in this specified range containing multiple data driven bandwidths for all data sets. This implies that any reasonable data driven bandwidth can work well here. As a representative illustration, Table 1 summarizes the averaged biases, SEs and MSEs of the estimates of β for bandwidth $h = 0.45$.

The three estimates of β_2 had very similar performance. This is consistent with the theory. Since $\underline{X}_2$ and $\underline{T}$ are independent and $\underline{X}_2$ is balanced among all subjects, it is expected theoretically that the three estimates should be equivalent at least up to the first order. The relative ratios of the absolute biases among three estimators ranged from 0.7 to 1.55. There was no one estimator consistently better than another.

For $\widehat{\beta}_1$, both the variance and absolute bias of the proposed estimate were uniformly smaller than that of the ZD estimate for each bandwidth considered, and the variances and absolute biases of both estimates were uniformly much smaller than those of the WI estimate. Compared to the WI estimate, the proposed and ZD estimates reduced the variances by more than 50%. The range of the absolute bias ratios of the WI estimate over the proposed estimate varied from 1.29 to 11.18. Nonetheless, Table 1 and a comparison between the variance and MSE plots in Figure 1 suggest that the bias is not of concern and the variance is a dominating factor when comparing the MSEs among the three estimates. Selecting the bandwidth equal to 0.45, the sandwich SE estimates of the proposed method agreed well with the empirical SEs. For example, when $h = 0.45$, the Monte Carlo SE of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ were 0.0567 and 0.1632 for the proposed method, while the averages of the corresponding sandwich estimated SEs were 0.0551 and 0.1612, respectively. Finally, the new nonparametric estimate

of $\theta(t)$ was more efficient than the WI estimate. The average MSEs of the WI estimates over the range of the bandwidths we considered were about 1.6 times of those using the proposed method.

7 Application to the Longitudinal CD4 Count Data

We applied the semiparametric model given in (1) and (2) to the longitudinal CD4 count data among HIV seroconverters previously analyzed by Zeger and Diggle (1994). This study involved 369 subjects whose CD4 counts were measured during a period of 3 years before to 6 years after seroconversion. A total of $2,376$ CD4 measurements were available and the number of CD4 observations per subject varied from 1 to 12, with the majority of subjects having 4 to 10 observations. It was of interest to estimate the average time course of CD4 counts and the effects of other covariates. These covariates included age, smoking status measured by packs of cigarettes, drug use with yes=1 and no=0, number of sex partners and depression status measured by the CESD Scale (large values indicating more depression symptoms). See Zeger and Diggle (1994) for a more detailed description of the data.

Let $\underline{T}$ be years since seroconversion. We conducted an analysis on the square-root-transformed CD4 counts using working independence and the proposed efficient estimator. The purpose behind the transformation is to reduce skewness of the original CD4 measurements, as indicated in Zeger and Diggle (1994). Our results in Section 6 indicate that neither estimator is sensitive to the choice of bandwidth. Therefore, we simply used a "partial" leaving-one-subject out cross-validation which dropped 50 randomly selected subjects one at a time to select a bandwidth of 1.86. To ensure our data analysis was indeed insensitive to the bandwidth selection, we repeated the analysis by reducing and increasing the bandwidth by 50%. The changes in coefficients and SEs were minimal. We hence report the results using the bandwidth 1.86.

For the proposed estimator, we used a working covariance structure described by ZD as "*random intercept plus serial correlation and measurement error*". More precisely, we assumed a random intercept and an exponential decay serial correlation by specifying the covariance structure as $\tau^2 I + \nu^2 J + \omega^2 H$, where J is a matrix of 1's and $H(j,k) = \exp(-\alpha|T_{ij} - T_{ik}|)$. The covariance estimates obtained by ZD were $\widehat{\xi} = (\hat{\tau}^2, \hat{\nu}^2, \hat{\omega}^2, \hat{\alpha}^2) = (14.1, 6.9, 16.1, 0.22)$. By leaving out residuals in the boundary and coupling a least square method in variogram analysis and a moment variance estimation approach, we obtained a slightly different set of estimates, $\widehat{\xi} = (11.32, 3.26, 22.15, 0.23)$. In Table 2, we referred to our and ZD's working covariances as "Scenario I" and "Scenario II" respectively.

Table 2 gives the regression coefficient estimates of the parametric covariates using the WI and proposed efficient method. The SEs were all calculated using the sandwich method. Based on the new method, smoking and the number of sex partners were significantly positively associated with the CD4 counts, while age, drug use and depression had no significant effects. The readers will note some fairly large numerical differences between the WI and the proposed estimates for smoking and drug use, a change of sign and statistical significance for number of sex partners and overall much smaller SEs for our method.

The decrease in SEs is in accordance with our theory. The other phenomena are more difficult to explain. Nonetheless, they are not unique to semiparametric GEE methods. Similar discrepant outcomes occurred in parametric GEE estimation in which $\theta(t)$ was replaced by a cubic regression function in time. Furthermore, we simulated data using the observed covariates but having responses generated from the multivariate normal with mean equal to the fitted mean in the parametric correlated GEE estimation, and with correlation given in Scenario II. The level of divergence between two sets of results in the simulated data was fairly consistent with what appeared in Table 2. For example, among the first 25 generated data sets, 3 had different signs in sex partners and 7 had the scale of drug use coefficient

obtained by WI 1.8 times or larger than what obtained by the proposed method. Among 100 generated datasets, the relationship between Monte Carlo estimated SEs and the Monte Carlo standard errors is basically the same as what obtained in parametric correlated GEE.

Comparing the estimates obtained in the two scenarios accounting for correlations, we note that using a slightly different covariance estimate does not change the outcome much even though the estimates are quite different with or without considering correlations.

The nonparametric curve estimates using the WI (dotted line) and proposed (solid line) estimators are plotted in the left panel of Figure 4. The CD4 counts were stable before seroconversion and sharply decreased after seroconversion. By accounting for correlation, our method suggests that the decreasing trend remained after 2 years. The estimated SEs are given in the right panel of Figure 4. The SE of the proposed curve estimate is uniformly smaller than that of the WI estimate. The results agree with the theory.

8 Discussion

We have considered the marginal semiparametric partial generalized linear model previously discussed in Lin and Carroll (2001a) for clustered data, where the effects of some covariates $\underline{X}$ are modeled parametrically and the effect of a covariate $\underline{T}$ is modeled nonparametrically as $\theta(t)$. LC showed the conventional profile-kernel method failed to yield a semiparametric efficient estimator of β. By simply replacing the nonparametric estimator in LC's original profile-kernel method by the newly proposed iterative kernel estimator $\widehat{\theta}(t)$, we were able to construct a semiparametric efficient estimator of β under the same multivariate normal scenario given in LC.

Unlike LC, a regular bandwidth can be used and under-smoothing is no longer needed to construct $\sqrt{n}$ consistent estimates of β when accounting for correlations. In addition, the proposed $\widehat{\theta}(t)$ has less variation than the WI estimator. Our numerical results suggest that

the proposed method performs well in finite sample and outperforms the WI method. They also suggests that the proposed $\widehat{\beta}$ is relatively insensitive to the choice of bandwidth.

Most importantly, we have shown that properly accounting for the within-subject correlation can reduce variation of parameter estimates in the general semiparametric model (2), just as in parametric models.

APPENDIX

Assume each f_j has a compact support and that on its support, f_j is bounded away from 0. Throughout the appendix, we assume the equivalent convexity conditions given in Carroll, et al. (1997) hold. These conditions ensure that the $\widehat{\alpha}$ and $\widehat{\beta}$ obtained in (3) and (4) exist uniquely and lie in a compact set. We also assume that conditions equivalent to Condition 2 of Carroll, et al. (1997) hold. The purpose behind these assumptions is to establish uniform convergence of $\widehat{\theta}$ and $\widehat{\varphi}$. The structure of the proof has been given in Mack and Silverman (1982) and the proof of Lemma A.1 and equation (A.5) in Carroll, et al. (1997). We further assume that $\int\int H^2(t,s)dtds < 1$. This condition assures the existence and uniqness of a solution to (6); see Kress (1989, chap. 2). With a linear link and i.i.d T_1, ..., T_m, this condition is equivalent to a constraint on dependent structure of the responses from the same subject. Except in the first half of Appendix A.1, we concentrate on $m_i \equiv m$. We let $n \to \infty$, $h \to 0$ at the rate that $nh^8 \to 0$ and $nh/\log(1/h) \to \infty$.

A.1 Semiparametric Efficient Score

Under the multivariate normal model, the joint density of $\{\underline{Y}_i, \underline{X}_i, \underline{T}_i\}$ is

$$f_{Y|X,T}(\underline{y}_1, \cdots, \underline{y}_n | \{\underline{x}_i, \underline{t}_i\}) f_{X,T}(\underline{x}_1, \cdots, \underline{x}_n, \underline{t}_1, \cdots, \underline{t}_n), \tag{A.1}$$

where $f_{X,T}$ is the joint density of $\{\underline{X}_i, \underline{T}_i\}$ and $f_{Y|X,T}$ is MVN with conditional means specified in (2) and conditional variance $\Sigma_i(\underline{X}_i, \underline{T}_i)$. Following Bickel, et al. (1993, chap. 3), we first define the following sub-models:

P_1: {Model (A.1) with known $\theta_0(\cdot)$ and $f_{X,T}(\cdot)$},

P_2: {Model (A.1) with known β_0 and $f_{X,T}(\cdot)$}, and

P_3: {Model (A.1) with known β_0 and $\theta_0(\cdot)$.}.

For the parametric family P_1, the score function for β_0 is

$$S_\beta = \sum_{i=1}^{n} \underline{X}_i^t \Delta_i \Sigma_i^{-1} [\underline{Y}_i - \mu\{\underline{X}_i\beta_0 + \theta_0(\underline{T}_i)\}].$$

By linearly spanning the score functions of parametric submodels of P_2 with $\theta(\cdot)$ replaced by $\theta(\eta, \cdot)$, the tangent space of P_2 is

$$\dot{P}_2 \;=\; \left\{ \sum_{i=1}^{n} \varphi^t(\underline{T}_i) \Delta_i \Sigma_i^{-1} [\underline{Y}_i - \mu\{\underline{X}_i\beta_0 + \theta_0(\underline{T}_i)\}],\ \text{where } \varphi(\cdot) \in L_2 \right\}.$$

We only need to concentrate on $\dot{P}_2$ because it is easy to see that S_β and any member in $\dot{P}_2$ are orthogonal to the score function in any parametric submodel of P_3; consequently, they are orthogonal to $\dot{P}_3$.

By Theorem 1 of Bickel, et al. (1993, Section 3.4), the efficient score $S_\beta^* = S_\beta - \prod(S_\beta|\dot{P}_2)$ is $\sum_i \{\underline{X}_i - \varphi_{eff}(\underline{T}_i)\}^t \Delta_i \Sigma_i^{-1} [\underline{Y}_i - \mu\{\underline{X}_i\beta_0 + \theta_0(\underline{T}_i)\}]$, where $\varphi_{eff}(\cdot)$ satisfies the requirement that S_β^* is orthogonal to any member in $\dot{P}_2$. That is, φ_{eff} needs to satisfy $\sum_i E\left[\{\underline{X}_i - \varphi_{eff}(\underline{T}_i)\}^t \Delta_i \Sigma_i^{-1} \Delta_i \varphi(\underline{T}_i)\right] = 0$, for any φ. This is equivalent to

$$n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_i} E\left[\Delta_{ijj} \sigma^{ij\ell} \Delta_{i\ell\ell} \left\{X_{i\ell} - \varphi_{eff}(T_{i\ell})\right\} \varphi(T_{ij})\right] = 0, \tag{A.2}$$

Up to this point, we have not assumed $m_i \equiv m$. The results in the main text are reported with this assumption for the purpose of a clean presentation. Without this assumption, but with a bound of the m_i, in the limit as $n \to \infty$, the efficient score (A.2) becomes the obvious weighted average of the interior sums $j, \ell = 1, ..., m_i$.

When $m_i \equiv m$ so that $(\underline{Y}_i, \underline{X}_i, \underline{T}_i)$ are i.i.d., (A.2) is equivalent to

$$\sum_{j=1}^{m} \sum_{\ell=1}^{m} E\left[\Delta_{jj} \sigma^{j\ell} \Delta_{\ell\ell} \left\{X_\ell - \varphi_{eff}(T_\ell)\right\} \varphi(T_j)\right] = 0,$$

which leads directly to (5).

A.2 Asymptotic Structure of $\widehat{\theta}$

A condition guaranteeing a unique solution to (6) is given at the beginning of the Appendix. The purpose of this section is to provide an asymptotic expansion for $\widehat{\theta}_{[k]}(t,\beta_0)$ at the kth iteration and at the convergence, when the iterations converge. As described in Section 3, the WI estimator is used as an initial estimator and is denoted by $\widehat{\theta}_{[0]}(t,\beta_0)$. Its asymptotic expansion is

$$\begin{aligned}\widehat{\theta}_{[0]}(t,\beta_0)-\theta(t) &= \frac{1}{2}b_{[0]}(t)h^2+W_2^{-1}(t)n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\mu_{ij}^{(1)}v_i^{jj}K_h(T_{ij}-t)(Y_{ij}-\mu_{ij})\\ &+ o_p(h^2+\{\log(n)/nh\}^{1/2}+n^{-1/2}),\end{aligned} \tag{A.3}$$

$$\text{where}\quad b_{[0]}(t)=\theta^{(2)}(t),\ \text{and}\ W_2(t)=\sum_{j=1}^{m}\mathrm{E}\{\Delta_{jj}^2v^{jj}|T_j=t\}f_j(t). \tag{A.4}$$

Recall that $f_{j\ell}(t,s)$ denotes the joint density of (T_j,T_ℓ) evaluated at (t,s). Define

$$Q(t,s) = \sum_j\sum_{\ell\neq j}E[\Delta_{jj}v^{j\ell}\Delta_{\ell\ell}\{W_2(T_\ell)\}^{-1}|T_j=t,T_\ell=s]f_{j\ell}(t,s); \tag{A.5}$$

$$b_{[k]}(t) = b_{[0]}(t)-W_2^{-1}(t)\sum_j\sum_{\ell\neq j}E\left\{\Delta_{jj}v^{j\ell}\Delta_{\ell\ell}b_{[k-1]}(T_\ell)|T_j=t\right\}f_j(t), \tag{A.6}$$

with $b_{[0]}(t)$ defined in (A.4). It can be shown that the estimated $\theta(\cdot)$ after one step update of $\widehat{\theta}_{[0]}$ has the following expansion:

$$\begin{aligned}\widehat{\theta}_{[1]}(t)-\theta(t) &= \frac{1}{2}b_{[1]}(t)h^2+W_2^{-1}(t)n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\mu_{ij}^{(1)}K_h(T_{ij}-t)\{\sum_{\ell=1}^{m}v_i^{jl}(Y_{i\ell}-\mu_{i\ell})\}\\ &- W_2^{-1}(t)n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\mu_{ij}^{(1)}v_i^{jj}Q(t,T_{ij})(Y_{ij}-\mu_{ij})+o_p(h^2+\{\log(n)/nh\}^{1/2}+n^{-1/2}).\end{aligned}$$

Further define an integration operator $\ddot{\mathcal{A}}\{B(\cdot,\cdot);t,s\}$:

$$\ddot{\mathcal{A}}(B;t,s)=-\sum_j\sum_{\ell\neq j}E[\Delta_{jj}v^{j\ell}\Delta_{\ell\ell}\{W_2(T_\ell)\}^{-1}B(T_\ell,s)|T_j=t]f_j(t). \tag{A.7}$$

For iteration $k\geq 2$, we have

$$\widehat{\theta}_{[k]}(t,\beta_0)-\theta(t) = \frac{1}{2}b_{[k]}(t)h^2+W_2^{-1}(t)n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\mu_{ij}^{(1)}K_h(T_{ij}-t)\{\sum_{\ell=1}^{m}v_i^{jl}(Y_{i\ell}-\mu_{i\ell})\}$$

$$
\begin{aligned}
&+ \quad W_2^{-1}(t) n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \mu_{ij}^{(1)} Q_{1,[k]}(t, T_{ij}) \{ \sum_{\ell=1}^{m} v_i^{jl} (Y_{i\ell} - \mu_{i\ell}) \} \\
&+ \quad W_2^{-1}(t) n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \mu_{ij}^{(1)} v_i^{jj} Q_{2,[k]}(t, T_{ij}) (Y_{ij} - \mu_{ij}) \\
&+ \quad o_p(h^2 + \{\log(n)/nh\}^{1/2} + n^{-1/2}), \qquad \text{(A.8)}
\end{aligned}
$$

where $b_{[k]}(t)$ is defined in (A.6), $Q_{1,[1]}(t,s) \equiv 0$; $Q_{2,[1]}(t,s) = -Q(t,s)$, and

$$
Q_{1,[k]}(t,s) = -Q(t,s) + \ddot{\mathcal{A}}(Q_{1,[k-1]}; t, s); \quad Q_{2,[k]}(t,s) = \ddot{\mathcal{A}}(Q_{2,[k-1]}; t, s).
$$

At convergence, $\widehat{\theta}_*(t) - \theta(t)$ shares the same asymptotic structure as in (A.8) except that $b_{[k]}$, $Q_{1,[k]}$ and $Q_{2,[k]}$ should be replaced by b_*, $Q_{1,*}$ and $Q_{2,*}$, respectively, where $\ddot{\mathcal{A}}$ is given in (A.7), b_*, $Q_{1,*}$ and $Q_{2,*}$ satisfy the corresponding integration equations:

$$
\begin{aligned}
b_*(t) &= \theta^{(2)}(t) - W_2^{-1}(t) \sum_j \sum_{\ell \neq j} E \left\{ \Delta_{jj} v^{j\ell} \Delta_{\ell\ell} b_*(T_\ell) | T_j = t \right\} f_j(t); \\
Q_{1,*}(t,s) &= -Q(t,s) + \ddot{\mathcal{A}}(Q_{1,*}; t, s); \quad Q_{2,*}(t,s) = \ddot{\mathcal{A}}(Q_{2,*}; t, s).
\end{aligned}
$$

A.3 Proof of Proposition 1

This appendix sketches a proof to show that in the Gaussian case assuming a linear link, $\varphi = \varphi_{eff}$, the latter given by (6) with Δ being the identity matrix. For the proof of the general case, simply place elements in Δ (the first derivative of μ) at the right places. At convergence, using (3), it can be shown that, for a given β, we have

$$
\begin{aligned}
0 \;=\; n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} K_h(t - T_{ij}) \Big(& \sigma_i^{jj} \left[Y_{ij} - X_{ij}\beta - \widehat{\theta}(t,\beta) - \{(T_{ij} - t)/h\} \widehat{\alpha}_1(t,\beta) \right] \\
& + \sum_{l \neq j} \sigma_i^{jl} \left\{ Y_{il} - X_{il}\beta - \widehat{\theta}(T_{il}, \beta) \right\} \Big).
\end{aligned}
$$

Taking derivatives with respect to β on both sides, direct derivations lead to

$$
\begin{aligned}
0 \;=\; n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} K_h(t - T_{ij}) \Big[& \sigma_i^{jj} \widehat{\varphi}(t) - \sigma_i^{jj} \{(T_{ij} - t)/h\} \, \partial \widehat{\alpha}_1(t,\beta)/\partial\beta \\
& - \sum_{\ell} \sigma_i^{j\ell} X_{il} + \sum_{\ell \neq j} \sigma_i^{j\ell} \widehat{\varphi}(T_{i\ell}) \Big].
\end{aligned}
$$

It is straightforward to show that $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}K_h(t-T_{ij})\sigma_i^{jj}\{(T_{ij}-t)/h\}=o_p(1)$, and

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\sigma_i^{jj}K_h(t-T_{ij}) = \left\{\sum_{j=1}^{m}E(\sigma^{jj}|T_j=t)f_j(t)\right\}\{1+o_p(1)\}, \quad (A.9)$$

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}K_h(t-T_{ij})\sum_{\ell=1}^{m}\sigma_i^{j\ell}X_{i\ell} = \left\{\sum_{j=1}^{m}\sum_{\ell=1}^{m}E\left(\sigma^{j\ell}X_\ell|T_j=t\right)f_j(t)\right\}\{1+o_p(1)\} \quad (A.10)$$

In addition, we have that

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}K_h(t-T_{ij})\{\sum_{\ell\neq j}\sigma^{j\ell}\widehat{\varphi}(T_{i\ell})\}=\sum_{j=1}^{m}\sum_{\ell\neq j}^{m}\int E(\sigma^{j\ell}|T_j=t)\widehat{\varphi}(t_\ell)f_{\ell j}(t_\ell,t)dt_\ell\{1+o_p(1)\} \quad (A.11)$$

The combinations of (A.9), (A.10), (A.11) lead to

$$\begin{aligned}&\sum_{j=1}^{m}E(\sigma^{jj}|T_j=t)f_j(t)\widehat{\varphi}(t)-\sum_{j=1}^{m}\sum_{\ell=1}^{m}E(\sigma^{j\ell}X_\ell|T_j=t)f_j(t)\\ -\ &\sum_{j=1}^{m}\sum_{\ell\neq j}^{m}\int E(\sigma^{j\ell}|T_j=t)\widehat{\varphi}(t_\ell)f_{\ell j}(t_\ell,t)dt_\ell=o_p(1), \quad (A.12)\end{aligned}$$

uniformly on t. Dividing (A.12) by $\sum_j E(\sigma^{jj}|T_j=t)f_j(t)$, noting that $\widehat{\varphi}(t)$ uniformly converges to $\varphi(t)$, and comparing the second and third terms to (7) and (8) with elements in $\Delta \equiv 1$, we can directly establish Proposition 1 by letting $\widehat{\varphi}(t)$ converges to $\varphi(t)$ in (A.12) and noting that $\varphi(t)$ fulfills (6).

A.4 Proof of (9)

The purpose of this section is to derive the asymptotic distribution of $\widehat{\beta}$. We first construct the following Lemma, which is a consequence of Proposition 1.

Lemma 1 *For any function $A(\cdot)$,*

$$\sum_j\sum_k E\left\{\widetilde{X}_j\Delta_{jj}v^{jk}\Delta_{kk}A(T_k)|T_k=t\right\}f_k(t)=0. \quad (A.13)$$

Furthermore,

$$\sum_j\sum_k E\left\{\widetilde{X}_j\Delta_{jj}v^{jk}\Delta_{kk}A(T_k)\right\}=0. \quad (A.14)$$

Proof of Lemma 1: We rewrite (5) by

$$\sum_j \sum_k \mathrm{E}\left\{\widetilde{X}_j \Delta_{jj} v^{jk} \Delta_{kk} | T_k = t\right\} f_k(t) = 0. \tag{A.15}$$

Equation (A.13) is established by multiplying both sides of (A.15) by $A(t)$ and noting that

$$\mathrm{E}\left\{\widetilde{X}_j \Delta_{jj} v^{jk} \Delta_{kk} A(T_k) | T_k = t\right\} = \mathrm{E}\left\{\widetilde{X}_j \Delta_{jj} v^{jk} \Delta_{kk} | T_k = t\right\} A(t).$$

Equation (A.14) follows directly from (A.13).

Proof of (9): Following the same derivations as in LC and keeping only the essential terms, some tedious calculations lead to the following asymptotic expansion for the profile estimator $\widehat{\beta}$:

$$n^{1/2}\{\widehat{\beta} - \beta\} = \left\{\widetilde{A}(V)\right\}^{-1} \{B_n + C_{1n} - C_{2n}\} + o_p(1),$$

where $\widetilde{A}(V)$ is defined in Proposition 2,

$$\begin{aligned}
B_n &= 1/2(nh^4)^{1/2} \left[n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{\ell=1}^{m} \mu_{ij}^{(1)} v_i^{j\ell} \mu_{i\ell}^{(1)} \widetilde{X}_{ij} \{b_*(T_{i\ell}) + h b_{*1}(T_{i\ell}) + O_p(h^2)\} \right] \{1 + o_p(1)\}; \\
C_{1n} &= n^{-1/2} \sum_{i=1}^{n} \widetilde{\underline{X}}_i^t \Delta_i(V_i)^{-1} (\underline{Y}_i - \underline{\mu}_i); \\
C_{2n} &= \left\{ n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{\ell=1}^{m} \widetilde{X}_{ij} \mu_{ij}^{(1)} v_i^{j\ell} \mu_{i\ell}^{(1)} \left(W_2^{-1}(T_{i\ell}) n^{-1} \sum_{i'=1}^{n} \sum_{j'=1}^{m} \mu_{i'j'}^{(1)} (v_{i'})^{j'j'} \right.\right. \\
&\quad \left[K_h(T_{i'j'} - T_{i\ell}) \left\{ \sum_l (v_{i'})^{j'l} (Y_{i'l} - \mu_{i'l}) \right\} + Q_{2,*}(T_{i\ell}, T_{i'j'}) (Y_{i'j'} - \mu_{i'j'}) \right. \\
&\quad \left.\left.\left. + \; Q_{1,*}(T_{i\ell}, T_{i'j'}) \left\{ \sum_l (v_{i'})^{j'l} (Y_{i'l} - \mu_{i'l}) \right\} \right] \right) \right\} \{1 + o_p(1)\},
\end{aligned}$$

with b_*, $Q_{1,*}$, and $Q_{2,*}$ being defined in Appendix A.2 and b_{*1} being the next order term in a higher order bias expansion of $\widehat{\theta}$ following the equivalent derivations in Theorem 1 of Fan, et al. (1995). In general, for an estimator of θ, e.g., the WI estimator, the square bracket term inside B_n is of order $O_p(1)$. Thus one needs nh^4 goes to 0 to eliminate the bias term in B_n. For the proposed estimator, as a consequence of (A.14), $B_n = o_p(1)$ provided that nh^8 goes to 0.

We now proceed to show that C_{2n} is of order $o_p(1)$. Write $C_{2n} = \sum_{\ell=1}^{3} C_{2\ell n}$, where

$$\begin{aligned} C_{21n} &= \frac{1}{\sqrt{n}} \sum_{i'=1}^{n} \sum_{j'=1}^{m} \mu_{i'j'}^{(1)} (v_{i'})^{j'j'} \left\{ n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{\ell=1}^{m} K_h(T_{i'j'} - T_{i\ell}) \widetilde{X}_{ij} \mu_{ij}^{(1)} v_i^{j\ell} \mu_{i\ell}^{(1)} W_2^{-1}(T_{i\ell}) \right\} \\ &\quad \left\{ \sum_{l} (v_{i'})^{j'l} (Y_{i'l} - \mu_{i'l}) \right\} \{1 + o_p(1)\}, \end{aligned}$$

where the term inside the first {} is asymptotically equivalent to

$$\sum_{j} \sum_{\ell} E \left\{ \widetilde{X}_j \Delta_{jj} v^{j\ell} \Delta_{\ell\ell} W_2^{-1}(T_\ell) | T_\ell = t \right\} f_\ell(t)|_{t=T_{i'j'}}. \tag{A.16}$$

Similarly, we have

$$\begin{aligned} C_{22n} &= \frac{1}{\sqrt{n}} \sum_{i'=1}^{n} \sum_{j'=1}^{m} \mu_{i'j'}^{(1)} (v_{i'})^{j'j'} \left[E \left\{ \widetilde{X}_j \Delta_{jj} v^{j\ell} \Delta_{\ell\ell} W_2^{-1}(T_\ell) Q_{2,*}(T_\ell, t) \right\} | t = T_{i'j'} \right] \\ &\quad (Y_{i'j'} - \mu_{i'j'}) \{1 + o_p(1)\}; \\ C_{23n} &= \frac{1}{\sqrt{n}} \sum_{i'=1}^{n} \sum_{j'=1}^{m} \mu_{i'j'}^{(1)} (v_{i'})^{j'j'} \left[E \left\{ \widetilde{X}_j \Delta_{jj} v^{j\ell} \Delta_{\ell\ell} W_2^{-1}(T_\ell) Q_{1,*}(T_\ell, t) \right\} | t = T_{i'j'} \right] \\ &\quad \left\{ \sum_{l} (v_{i'})^{j'l} (Y_{i'l} - \mu_{i'l}) \right\} \{1 + o_p(1)\}. \end{aligned}$$

By Lemma 1, we observe that the expectation terms in (A.16), C_{22n} and C_{23n} all equal zero. They are also only functions of T and X. Thus, with $h \to 0$ and $n \to \infty$ at the rates that $nh^8 \to 0$ and $nh/\log(1/h) \to \infty$, we obtain that B_n, C_{21n} and C_{22n} are all of order $o_p(1)$.

It follows that

$$n^{1/2}(\widehat{\beta} - \beta) = \left\{ \widetilde{A}(V) \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underline{\widetilde{X}}_i^t \Delta_i V_i^{-1} (\underline{Y}_i - \underline{\mu}_i) \{1 + o_p(1)\}. \tag{A.17}$$

which implies that $n^{1/2}(\widehat{\beta} - \beta) \to \text{Normal}\{0, \Omega(V, \Sigma)\}$, where $\Omega(V, \Sigma)$ is defined in Proposition 2. An application of an extended Cauchy-Schwartz inequality given in Johnson and Wichern (1982, §2.7) indicates that the best choice of V is Σ, which gives $\Omega(V, \Sigma) = \widetilde{A}^{-1}(\Sigma)$.

A.5 Derivation of (10) and Structure of $\underline{Z}_1$ and $\underline{Z}_2$

Under the normality assumption of $\underline{X}$ and $\underline{T}$, $E(X_\ell | T_j = t) = (\sigma_X \rho_{XT} / \sigma_T) t$. This equation and the structure of (6) imply that $\varphi_{eff}(t) = c(\sigma_X / \sigma_T) t$, where c is some constant. Plugging

this into (6), it is straightforward to solve for c and obtain (10). Similar calculations can be used to obtain $I_{WI}(\beta)$ and $I_{ZD}(\beta)$ given in Section 5, where using Result 1 of LC, $\underline{Z}_1$ and $\underline{Z}_2$ in $I_{ZD}(\beta)$ are $\underline{Z}_1 = \Sigma^{-1}\widetilde{\underline{X}}_{WI}$ and the k-th row of $\underline{Z}_2$ is

$$\frac{\sigma_{kk}^{-1}\sum_{j=1}^{m}\sum_{\ell=1}^{m}\sigma^{j\ell}\mathrm{E}(\widetilde{\underline{X}}_{WI,j}|T_\ell = T_k)}{\sum_{j=1}^{m}\sigma_{jj}^{-1}}.$$

REFERENCES

Begun, J. H., Hall, W. J., Huang, W. M. and Wellner, J. A. (1983), "Information and Asymptotic Efficiency in Parametric–Nonparametric Models", *Ann. Statist.*, 11, 432–452.

Bickel, P. J., Klaassen, A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Inference in Semiparametric Models,* Baltimore: Johns Hopkins University Press.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997), "Generalized Partially Linear Single-Index Models," *J. Am. Statist. Assoc.*, 92, 477-489.

Carroll, R. J., Wu, C. F. J. and Ruppert, D. (1988), "The Effect of Estimating Weights in Linear Regression," *J. Am. Statist. Assoc.*, 83, 1045-1054.

Crowder, M. (2001), "On Repeated Measures Analysis with Misspecified Covariance Structure," *J. Royal Statist. Soc., Ser. B*, 63, 55-62.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994), *Analysis of Longitudinal Data*, Oxford: Oxford University Press.

Fan, J and Gijbels, I., Hu, T. C. and Huang, L. S. (1995), "A Study of Variable Bandwidth Selection for Local Polynomial Regression," *Statist. Sinica* , 6, 113-127.

Heagerty, P. J. and Zeger, S. L. (2000), "Marginalized Multilevel Models and Likelihood Inference," *Statist. Sci.*, 15, 1-26.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, Y. (1998), "Nonparametric Smoothing Estimates of Time–Varying Coefficient Models with Longitudinal Data," *Biometrika*, 85, 809–822.

Johnson, R. A. and Wichern, D. W. (1982), *Applied Multivariate Statistical Analysis*, New York: Prentice and Hall.

Kress, R. (1989), *Linear Integral Equations,* Berlin Heidelberg: Springer-Verlag.

Lin, X. and Carroll, R. J. (2001a), "Semiparametric Regression for Clustered Data Using Generalized Estimating Equations," *J. Am. Statist. Assoc.*, 96, 1045-1056.

Lin, X. and Carroll, R. J. (2001b), Semiparametric Regression for Clustered Data. *Biometrika*, 88, 1179-1865.

Lin, X., Wang, N., Welsh, A. and Carroll, R. J. (2003) Equivalent Kernels of Smoothing Splines in Nonparametric Regression for Clustered Data. preprint.

Lin, D. Y. and Ying, Z. (2001), "Semiparametric and Nonparametric Regression Analysis of Longitudinal Data (With Discussion)," *J. Am. Statist. Assoc.*, 96, 103–126.

Mack, Y. and Silverman, B. (1982), "Weak and Strong Uniform Consistency of Kernel Regression Estimates," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete,* 61, 405-415.

Pepe, M. S. and Couper, D. (1997), "Modeling Partly Conditional Means with Longitudinal Data," *J. Am. Statist. Assoc.*, 92, 991- 998.

Prentice, R. and Zhao, L. P. (1991), "Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses," *Biometrics*, 47, 825-839.

Ruppert, D. (1997), "Empirical-bias Bandwidths for Local Polynomial Nonparametric Regression and Density Estimation," *J. Am. Statist. Assoc.*, 92, 1049-1062.

Ruppert, D. and Wand, M. P. (1994), "Multivariate Weighted Least Squares Regression," *Ann. Statist.*, 22, 1346-1370.

Severini, T. A. and Staniswalis, J. G. (1994), "Quasilikelihood Estimation in Semiparametric Models," *J. Am. Statist. Assoc.*, 89, 501–511.

Wang, N. (2003), "Marginal Nonparametric Kernel Regression Accounting for Within-Subject Correlation," *Biometrika*, 43-52..

Wild, C. J. and Yee, T. W. (1996). "Additive Extensions to Generalized Estimating Equation Methods," *J. Royal Statist. Soc., Ser. B*, 58, 711-725.

Zeger, S. L. and Diggle, P. J. (1994), "Semi-parametric Models for Longitudinal Data With Application to CD4 Cell Numbers in HIV Seroconverters," *Biometrics*, 50, 689-699.

Table 1: *Summary of simulation study results from 250 replications. The bandwidth used is* 0.45. *Each entry equals the original value multiplied by* 10.

Parameter	Method	Bias	SE	MSE
$\beta_1 = 1$	WI	.0732	.8564	.0739
	ZD	.0222	.5803	.0337
	NEW	.0118	.5675	.0322
$\beta_2 = 1$	WI	.0135	1.6486	.2718
	ZD	.0134	1.6379	.2683
	NEW	.0107	1.6324	.2665

Table 2: *Regression Coefficients in the CD4 cell counts study in HIV seroconverters using the Semiparametric Efficient and the Working Independence Estimate. For the semiparametric efficient estimates, the working covariance parameter,* $\widehat{\xi}$ =$(11.32, 3.26, 22.15, 0.23)$ *for Scenario I, and* $\widehat{\xi}$ = *(*$14.1, 6.9, 16.1, 0.22$*), for Scenario II.*

	Working Independence		Semiparametric Efficient Scenario I		Semiparametric Efficient Scenario II	
	Estimate	SE	Estimate	SE	Estimate	SE
Age	.014	.035	.010	.033	.008	.032
Smoking	.984	.182	.549	.144	.579	.139
Drug	1.049	.526	.584	.331	.584	.335
Sex Partners	- .054	.059	.080	.038	.078	.039
Depression	- .033	.021	-.045	.013	-.046	.014

LIST OF ILLUSTRATIONS

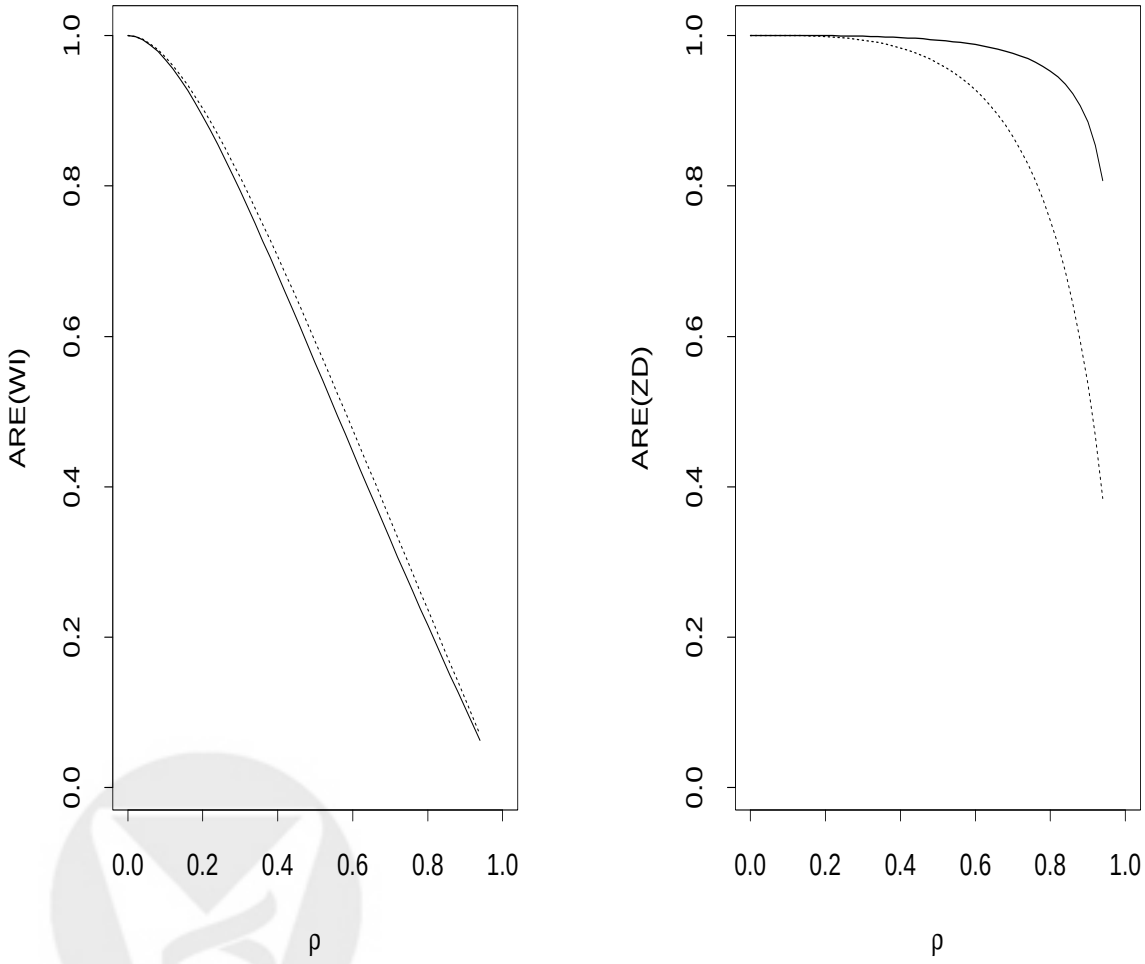

Figure 1:

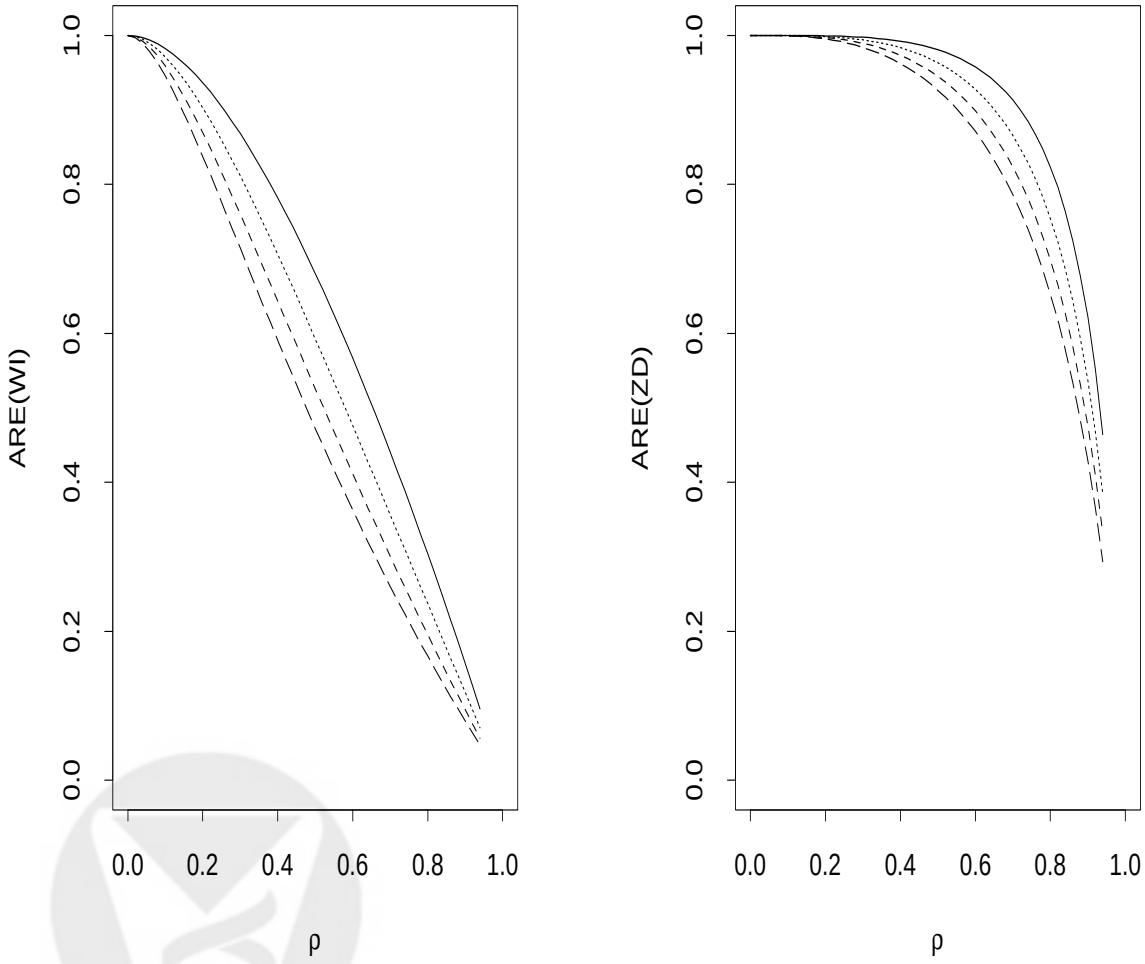

Figure 2:

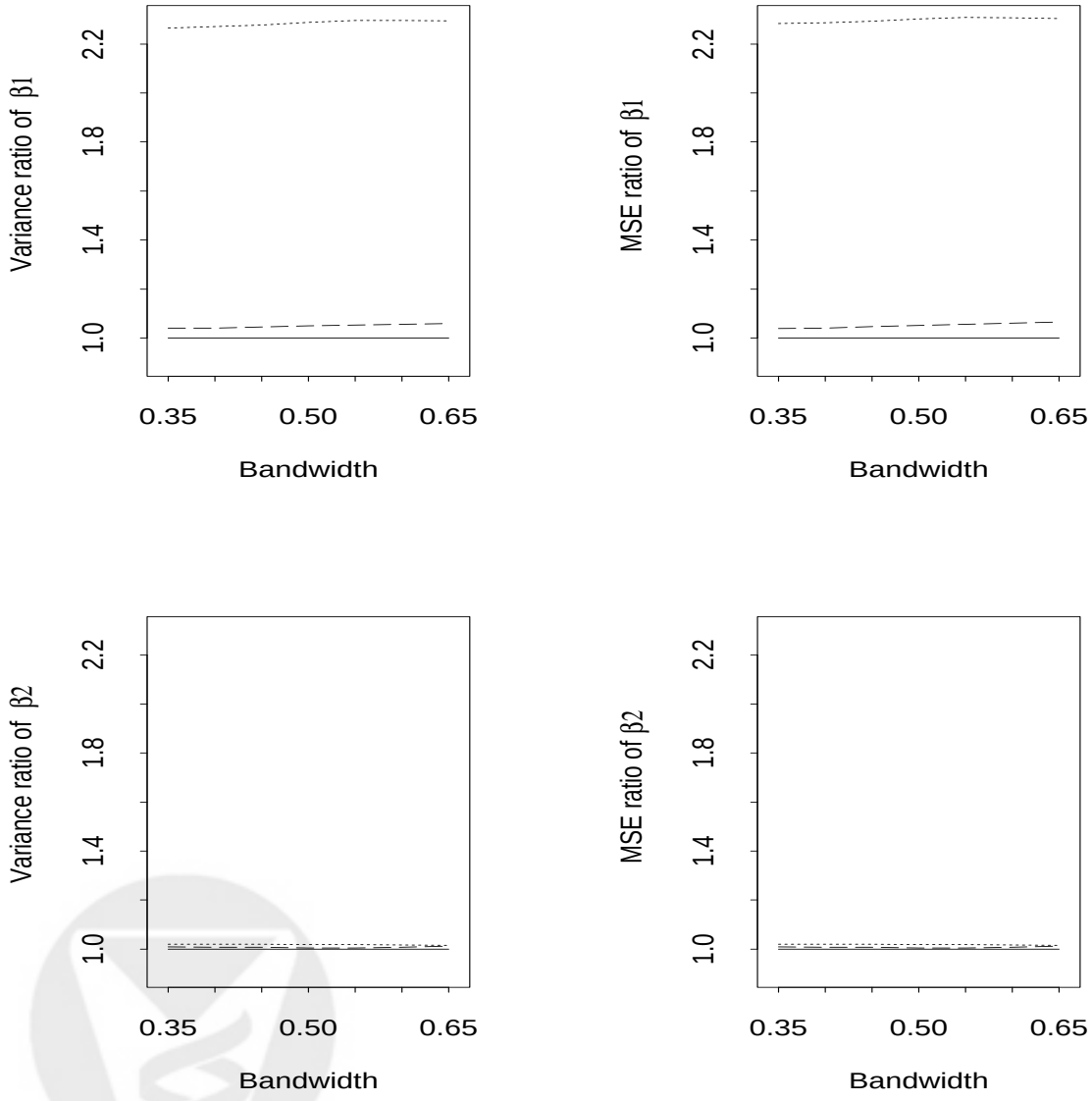

Figure 3:

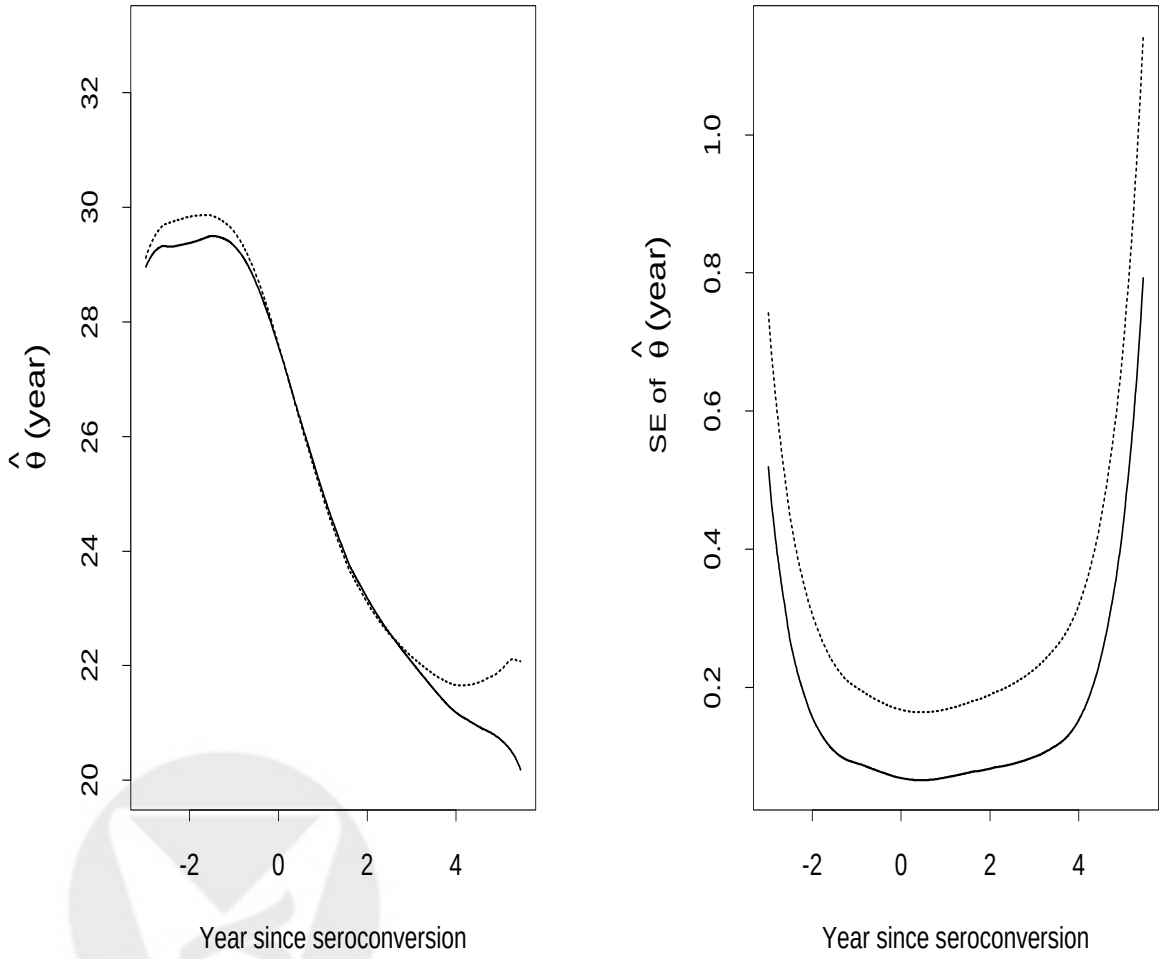

Figure 4: