Published in final edited form as:

Stat Biosci. 2009 May 1; 1(1): 10–31. doi:10.1007/s12561-009-9000-7.

Efficient Semiparametric Marginal Estimation for the Partially Linear Additive Model for Longitudinal/Clustered Data

Raymond Carroll,

Department of Statistics, 3143 TAMU, Texas A&M University, College Station, Texas 77843, USA, carroll@stat.tamu.edu, Telephone 979 845 3141, Fax 979 845 3144

Arnab Maity,

Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, U.S.A. amaity@hsph.harvard.edu

Enno Mammen, and

Department of Economics, University of Mannheim, L 7, 3–5, 68131 Mannheim, Germany, emammen@rumms.uni-mannheim.de

Kyusang Yu

Department of Economics, University of Mannheim, L 7, 3-5, 68131 Mannheim, Germany, kyusangu@yahoo.co.kr

Abstract

We consider the efficient estimation of a regression parameter in a partially linear additive nonparametric regression model from repeated measures data when the covariates are multivariate. To date, while there is some literature in the scalar covariate case, the problem has not been addressed in the multivariate additive model case. Ours represents a first contribution in this direction. As part of this work, we first describe the behavior of nonparametric estimators for additive models with repeated measures when the underlying model is not additive. These results are critical when one considers variants of the basic additive model. We apply them to the partially linear additive repeated-measures model, deriving an explicit consistent estimator of the parametric component; if the errors are in addition Gaussian, the estimator is semiparametric efficient. We also apply our basic methods to a unique testing problem that arises in genetic epidemiology; in combination with a projection argument we develop an efficient and easily computed testing scheme. Simulations and an empirical example from nutritional epidemiology illustrate our methods.

Keywords

Additive models; Generalized least squares; Interaction testing; Nonparametric regression; Partially linear model; Repeated measures; Smooth backfitting; Tukey-type models

1 Introduction

We consider the efficient estimation of a regression parameter in a partially linear additive nonparametric regression model from repeated measures data when the covariates are multivariate. To date, while there is some literature in the scalar covariate case, see below, the problem has not been addressed in the multivariate additive model case. Ours represents a first contribution in this direction.

There has been considerable interest in the simplest version of this problem. Suppose that there are $i = 1, \dots, n$ individuals, and $j = 1, \dots, J$ observations per individual. The responses are

Y_{ij} , the scalar predictors modeled nonparametrically are X_{ij} , and the possibly vector-values parameters modeled parametrically are Z_{ij} . A simple model says that given $(X_{i1}, \dots, X_{iJ})$,

$$\begin{aligned} Y_{ij} &= Z_{ij}^\top \theta_{\text{true}} + m_{\text{true}}(X_{ij}) + \epsilon_{ij}; \\ \text{cov}(\tilde{\epsilon}_i) &= \text{cov}(\epsilon_{i1}, \dots, \epsilon_{iJ})^\top = \Sigma_{\text{true}}, \end{aligned} \quad (1)$$

where $\tilde{\epsilon}$ has mean zero. This is the classical partially linear model in the repeated measures case.

When there are no repeated measures ($J = 1$), there is a vast literature on estimating θ_{true} under various conditions (Härdle et al. 2000). With the repeated measures, estimation using kernel methods has been treated (Wang et al. 2005) and considerably generalized (Lin and Carroll 2006).

In this paper, our goal is to consider repeated measures models such as (1) in the case that the argument of $m_{\text{true}}(\cdot)$ is multivariate, rather than a scalar. As is usual, to avoid the curse of dimensionality we take an additive modeling approach. Suppose we have a random sample $\{(Y_i^\top, \mathbf{X}_i^\top)^\top\}$ for $i = 1, \dots, n$, where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})^\top$ and $\mathbf{X}_i = (X_{i11}, \dots, X_{i1J}, \dots, X_{iD1}, \dots, X_{iDJ})^\top$ with

$$\begin{aligned} Y_{ij} &= Z_{ij}^\top \theta_{\text{true}} + m_{0,\text{true}} + \sum_{d=1}^D m_{d,\text{true}}(X_{idj}) + \epsilon_{ij}; \\ \tilde{\epsilon}_i &= (\epsilon_{i1}, \dots, \epsilon_{iJ})^\top \sim (\mathbf{0}, \Sigma_{\text{true}}), \end{aligned} \quad (2)$$

where Σ_{true} has elements $\sigma_{\ell k, \text{true}}$ and is positive definite. We have J repeated measurements consisting of one response and D regressors from n individuals. Assume $\mathbf{X}_1$ has the density $p(\cdot)$. We denote the density of $(X_{11\ell}, \dots, X_{1D\ell})^\top$ as $p_{\ell\ell}(\cdot)$, the density of $(X_{1sk}, X_{1b\ell})^\top$ as p_{kl}^{sb} and the density of X_{1dk} as p_k^d .

As far as we are aware, there has been no considerations of efficient estimation of θ_{true} in model (2), and it is the purpose of this paper to provide such estimates and the accompanying theory. If θ_{true} were known, then the functions $\{m_{d,\text{true}}(\cdot)\}$ can be estimated efficiently by a local linear smooth backfitting algorithm (Carroll et al. 2009) and reviewed in Section 2. We combine this methodology with a semiparametric profiling argument (Lin and Carroll 2006) to develop estimates of θ_{true} and inference for it in Section 3.

While the algorithm for estimating θ_{true} is quite simple, technically the development is challenging because it *involves fitting additive nonparametric models to data that do not have the additive structure*. Such a technical issue does not arise in simpler models such as (1), where there is only one nonparametric function being estimated. In particular, as part of the work, we describe the properties of additive model regression when the additive model fails, a result of interest in itself.

Our paper is organized as follows. In Section 2, we review the basic repeated measures smooth backfitting algorithm for local linear regression. Included in this section is a new result about the properties of additive model regression when the additive model fails. Section 3 gives the algorithm for estimating θ_{true} , and states the asymptotic results for it. A simulation study is described in Section 4, while Section 5 describes an empirical data set from nutritional epidemiology. Section 6 describes a non-standard model arising in genetics in which it is interesting to test a null hypothesis that a parameter equals zero. We develop a score test approach for this problem, the theory for which again uses our new theory of misspecified

additive models. Section 7 gives concluding remarks. Proofs are sketched in a technical appendix.

2 Local Linear Smooth Backfitting and its Properties

In this section, we describe the local linear smooth backfitting estimator for repeated measures data (Carroll et al. 2009), state its properties when an additive model holds, and then describe its properties when the additive model fails. The latter result is new.

2.1 The Local Linear Estimator

Let K^0 be a base kernel function and $K_h^0(u) = h^{-1}K^0(u/h)$. Define a boundary corrected kernel function by $K_h(u, v) = I(u, v \in S)K_h^0(u - v) / \int_S K_h^0(w - v)dw$, where S is the support.

We allow for estimation of the covariance matrix Σ_{true} . Let the elements of $\Sigma_{\text{true}}^{-1}$ be $b_{jk, \text{true}}$, and for an arbitrary covariance matrix Σ , let the elements of Σ^{-1} be $b_{jk, \Sigma}$. Write the symmetric matrix $\Sigma^{-1/2}$ to have elements $a_{j\ell, \Sigma}$. Note that $\sum_{\ell=1}^J a_{j\ell, \Sigma} a_{k\ell, \Sigma} = b_{jk, \Sigma}$. One choice of Σ is the covariance estimator based on the residuals of the pooled data, but more complex generalizations are easily handled.

Let $\mathbf{1}$ be a vector of ones and $\widehat{m}_0^\Sigma = n^{-1} \sum_{i=1}^n \mathbf{1}^\top \Sigma^{-1} \mathbf{Y}_i / \mathbf{1}^\top \Sigma^{-1} \mathbf{1}$. For a local linear fit, consider the smoothed sum of squares given as follows:

$$S(m_1, \dots, m_D, m^1, \dots, m^D) = \int n^{-1} \sum_{i=1}^n \left[\sum_{j,k=1}^J b_{jk, \Sigma} \{Y_{ij} - \widehat{m}_0^\Sigma - \sum_{d=1}^D m_d(x_{dj}) - \sum_{d=1}^D \frac{X_{idj} - x_{dj}}{h_d} m^d(x_{dj})\} \times \{Y_{ik} - \widehat{m}_0^\Sigma - \sum_{d=1}^D m_d(x_{dk}) - \sum_{d=1}^D \frac{X_{idk} - x_{dk}}{h_d} m^d(x_{dk})\} \right] \times \prod_{k=1}^J K_{h_1}(x_{1k}, X_{i1k}) \cdots K_{h_D}(x_{Dk}, X_{iDk}) d\mathbf{x}. \quad (3)$$

Here, $\mathbf{x} = (x_{11}, \dots, x_{1J}, \dots, x_{D1}, \dots, x_{DJ})^\top$. The goal of smooth backfitting is to minimize (3) in the functions $(m_1, \dots, m_D, m^1, \dots, m^D)$.

Define the following density estimators

$\widehat{p}_k^a(x) = n^{-1} \sum_{i=1}^n K_{h_a}(x, X_{iak})$, $\widehat{p}_{jk}^{ab}(x, y) = n^{-1} \sum_{i=1}^n K_{h_a}(x, X_{iaj}) K_{h_b}(y, X_{ibk})$ for $j, k = 1, \dots, J$ and $a, b = 1, \dots, D$. The minimizer of (3) exists and is given as follows (Carroll et al 2009). Define the following estimators, for $q = 1, \dots, D$, $r = 0, 1, 2$,

$$\widehat{u}_s^r(x) = \sum_{j=1}^J b_{jj, \Sigma} n^{-1} \sum_{i=1}^n \left(\frac{X_{iqj} - x}{h_q} \right)^r K_{h_q}(x, X_{iqj})$$

and for $q, p = 1, \dots, D$; $r, s = 0, 1$,

$$\begin{aligned} \widehat{v}_q^{rs}(x, y) &= \sum_{j=1}^J \sum_{k=1, k \neq j}^J b_{jk, \Sigma} n^{-1} \sum_{i=1}^n \left(\frac{X_{iqj} - x}{h_q} \right)^r \left(\frac{X_{ipk} - y}{h_p} \right)^s K_{h_q}(x, X_{iqj}) K_{h_p}(y, X_{ipk}); \\ \widehat{v}_{qp}^{rs}(x, y) &= \sum_{j=1}^J \sum_{k=1}^J b_{jk, \Sigma} n^{-1} \sum_{i=1}^n \left(\frac{X_{iqj} - x}{h_q} \right)^r \left(\frac{X_{ipk} - y}{h_p} \right)^s K_{h_q}(x, X_{iqj}) K_{h_p}(y, X_{ipk}). \end{aligned}$$

Then we obtain the following system of fitting integral equations: for $d = 1, \dots, D$,

$$\begin{aligned} \begin{Bmatrix} m_d(x_d) \\ \tilde{m}^d(x_d) \end{Bmatrix} &= \begin{Bmatrix} \tilde{m}_d^\Sigma(x_d) \\ \tilde{m}^{d,\Sigma}(x_d) \end{Bmatrix} - \widehat{\mathbf{M}}_d^{-1}(x_d) \int \widehat{\mathbf{V}}_d(x_d, t) \begin{Bmatrix} m_d(t) \\ m^d(t) \end{Bmatrix} dt \\ &\quad - \sum_{s=1, \neq d}^D \widehat{\mathbf{M}}_d^{-1}(x_d) \int \widehat{\mathbf{V}}_{ds}(x_d, t) \begin{Bmatrix} m_s(t) \\ m^s(t) \end{Bmatrix} dt, \end{aligned} \quad (4)$$

where

$$\begin{aligned} \widehat{\mathbf{M}}_d(x_d) &= \begin{Bmatrix} \widehat{u}_d^0(x_d) & \widehat{u}_d^1(x_d) \\ \widehat{u}_d^1(x_d) & \widehat{u}_d^2(x_d) \end{Bmatrix}; \\ \begin{Bmatrix} \tilde{m}_d^\Sigma(x_d) \\ \tilde{m}^{d,\Sigma}(x_d) \end{Bmatrix} &= \widehat{\mathbf{M}}_d^{-1}(x_d) \begin{Bmatrix} n^{-1} \sum_{i=1}^n \sum_{j,k=1}^J b_{jk,\Sigma}(Y_{ik} - \widehat{m}_0^\Sigma) K_{h_d}(x, X_{idj}) \\ n^{-1} \sum_{i=1}^n \sum_{j,k=1}^J b_{jk,\Sigma}(Y_{ik} - \widehat{m}_0^\Sigma) \frac{X_{idj} - x}{h_d} K_{h_d}(x, X_{idj}) \end{Bmatrix}; \\ \widehat{\mathbf{V}}_d(x_d, t) &= \begin{Bmatrix} \widehat{v}_d^{0,0}(x_d, t) & \widehat{v}_d^{0,1}(x_d, t) \\ \widehat{v}_d^{1,0}(x_d, t) & \widehat{v}_d^{1,1}(x_d, t) \end{Bmatrix}; \widehat{\mathbf{V}}_{ds}(x_d, t) = \begin{Bmatrix} \widehat{v}_{ds}^{0,0}(x_d, t) & \widehat{v}_{ds}^{0,1}(x_d, t) \\ \widehat{v}_{ds}^{1,0}(x_d, t) & \widehat{v}_{ds}^{1,1}(x_d, t) \end{Bmatrix}, \end{aligned} \quad (5)$$

for $d, s = 1, \dots, D$. They impose the identification condition

$$\int \sum_{k=1}^J \sum_{j=1}^J b_{jk,\Sigma} \left\{ m_d(x) \widehat{p}_j^d(x) + m^d(x) n^{-1} \sum_{i=1}^n \left(\frac{X_{idj} - x}{h_d} \right) K_{h_d}(x, X_{idj}) \right\} dx = 0. \quad (6)$$

It is easy to check that the solution of Equation (4) automatically satisfies the identification condition.

A backfitting algorithm (Carroll et al. 2009) can be used to implement this estimator without having to compute the high dimensional integrals in (3). Define

$$\begin{aligned} \{\widehat{\mathcal{G}}_{dd}^\Sigma(f, g)^\top\}(x) &= \widehat{\mathbf{M}}_d^{-1}(x) \int \widehat{\mathbf{V}}_d(x, t) \{f(t), g(t)\}^\top dt; \\ \{\widehat{\mathcal{G}}_{da}^\Sigma(f, g)^\top\}(x) &= \widehat{\mathbf{M}}_d^{-1}(x) \int \widehat{\mathbf{V}}_{da}(x, t) \{f(t), g(t)\}^\top dt. \end{aligned}$$

Then the backfitting algorithm takes the following form. Equation (4) can be rewritten as

$$\{(I + \widehat{\mathcal{G}}_{dd}^\Sigma)(m_d, m^d)^\top\}(x_d) = \begin{Bmatrix} \tilde{m}_d^\Sigma(x_d) \\ \tilde{m}^{d,\Sigma}(x_d) \end{Bmatrix} - \sum_{s=1, \neq d}^D \{\widehat{\mathcal{G}}_{ds}(\widehat{m}_s, \widehat{m}^s)^\top\}(x_d), \quad d=1, \dots, D,$$

where I is the identity mapping. Based upon this observation, we build up a backfitting type algorithm, as follows.

1. Set the initial values $m_d^{[0]}$ and $m^{d[0]}$ for $d = 1, \dots, D$ which satisfy the identification condition.
2. Let $(m_d^{[k]}, m^{d[k]})$ be the estimates of the d^{th} function in the k^{th} iteration. The updating equation is defined as

$$\begin{aligned} \{(I + \widehat{\mathcal{G}}_{dd}^\Sigma)(m_d^{[k]}, m^{d[k]})^\top\}(x_d) &= \begin{Bmatrix} \tilde{m}_d^\Sigma(x_d) \\ \tilde{m}^{d,\Sigma}(x_d) \end{Bmatrix} (x_d) \\ &\quad - \sum_{s=1}^{d-1} \{\widehat{\mathcal{G}}_{ds}(m_s^{[k]}, m^{s[k]})^\top\}(x_d) - \sum_{s=d+1}^D \{\widehat{\mathcal{G}}_{ds}(m_s^{[k-1]}, m^{s[k-1]})^\top\}(x_d). \end{aligned}$$

Note that in practice the integrations in the operators and the inversion of the operators $I+\mathcal{G}_{dd}^{\Sigma}$ can be done by discrete approximations of functions on grid points. In our simulations and in our data example we applied the trapezoidal rule for numerical integration.

3. Iterate this updating until it converges. Convergence of the algorithm has been proven (Carroll et al. 2009).

2.2 Properties When the Additive Model Holds

The following results are known (Carroll et al. 2009). Make the following assumptions.

ASSUMPTIONS

- A1. For $j, k = 1, \dots, J$ and $d, s = 1, \dots, D$, p_{jk}^{ds} is bounded away from zero and infinity on its support, $[0, 1]^2$, and have continuous partial derivatives.
- A2. $E|Y_{1j}|^{r_0} < \infty$ for $j = 1, \dots, J$ and some $r_0 > 5/2$.
- A3. The true component functions $m_{d,\text{true}}$'s are twice continuously differentiable.
- A4. The base kernel function K^0 is a symmetric density function with compact support, $[-1, 1]$ say, and is Lipschitz continuous.
- A5. $n^{1/5}h_d$ converge to constants $c_d > 0$ for $d = 1, \dots, D$ as n goes to infinity.

Let Σ be any positive definite matrix. We say $\hat{\Sigma}$ converges to Σ in probability if

$$\sup_{x: \|x\|_2=1} x^\top (\hat{\Sigma} - \Sigma)x = o_p(1).$$

Let $P^d(x_d) = \text{diag}\{p_1^d(x_d), \dots, p_J^d(x_d)\}$. Define, for $d = 1, \dots, D$,

$$\begin{aligned} v_d^\Sigma(x_d) &= c_d^{-1} \mu_0 \{ (K^0)^2 \} \text{trace} \{ \Sigma^{-1} \Sigma_{\text{true}} \Sigma^{-1} P^d(x_d) \} \left\{ \sum_{j=1}^J b_{jj, \Sigma} p_j^d(x_d) \right\}^{-2}; \\ \beta_d^{LL}(x_d) &= (1/2) \mu_2(K^0) h_d^2 m_{d,\text{true}}''(x_d), \end{aligned}$$

where $\mu_r(K) = \int t^r K(t) dt$ for a function K . They show that if the positive definite estimator $\hat{\Sigma}$ converges to Σ in probability, it follows that for any $x_1, \dots, x_d \in (0; 1)$,

$$\begin{aligned} n^{2/5} \begin{Bmatrix} \hat{m}_1^{\hat{\Sigma}}(x_1) - m_{1,\text{true}}(x_1) - \beta_1^{LL}(x_1) \\ \vdots \\ \hat{m}_D^{\hat{\Sigma}}(x_D) - m_{D,\text{true}}(x_D) - \beta_D^{LL}(x_D) \end{Bmatrix} &\rightarrow \text{Normal} \left[\mathbf{0}, \text{diag}\{v_1^\Sigma(x_1), \dots, v_D^\Sigma(x_D)\} \right]; \\ n^{1/2} (\hat{m}_0^{\hat{\Sigma}} - m_{0,\text{true}}) &\rightarrow \text{Normal} \left\{ 0, \frac{\mathbf{1}^\top \Sigma^{-1} \Sigma_{\text{true}} \Sigma^{-1} \mathbf{1}}{(\mathbf{1}^\top \Sigma^{-1} \mathbf{1})^2} \right\}. \end{aligned}$$

If the covariance estimator $\hat{\Sigma}$ converges to Σ_{true} in probability, for $d = 1, \dots, D$, we have $\beta_{d,\text{true}}(x_d)$ defined with Σ_{true} instead of Σ and

$$\begin{aligned} u_{d,\text{true}}(x_d) &= c_d^{-1} \mu_0 \{ (K^0)^2 \} \left\{ \sum_{\ell=1}^J b_{\ell\ell,\text{true}} p_{\ell}^d(x_d) \right\}^{-1}; \\ n^{1/2}(\widehat{m}_0^{\Sigma} - m_{0,\text{true}}) &\rightarrow \text{Normal} \left\{ 0, (\mathbf{1}^{\top} \Sigma_{\text{true}}^{-1} \mathbf{1})^{-1} \right\}. \end{aligned}$$

Remark 1: The bias term of our estimators is the same as for ordinary nonparametric regression with no repeated measures and no additive model. This bias term in our estimator does not involve the density function of the X 's, which is called design independent bias (Fan and Gilbels 1996). Also, when the covariance estimator $\hat{\Sigma}$ converges to Σ_{true} in probability, our estimator is efficient.

Remark 2: The assumption A5 means that ordinary bandwidth estimation will work in our context, and the semiparametric contexts to follow. Section 3.2 of Carroll et al. (2009) discusses this issue. Also see Maity, et al. (2007) for other comments on bandwidth estimation in semiparametric models.

2.3 When the True Regression Function is not Additive

In this section, we do not assume an additive model but instead assume only that

$$\begin{aligned} Y_{ij} &= m(X_{i1j}, \dots, X_{iDj}) + \epsilon_{ij}; \\ \text{cov}(\tilde{\epsilon}_i) &= \text{cov}(\epsilon_{i1}, \dots, \epsilon_{iJ})^{\top} = \Sigma_{\text{true}}, \end{aligned} \quad (7)$$

and where $\tilde{\epsilon}$ has mean zero.

One can expect that the proposed smooth backfitting type estimator converges to an additive projection of $m(\cdot)$ under the model (7), because the estimator minimizes a distance between responses and additive functions. We denote this additive projection by

$m^{\text{add}}(x_1, \dots, x_D) = m_0^* + \sum_{d=1}^D m_d^*(x_d)$. The explicit definition of the m_d^* 's is given in equation (A.2) of Appendix A.1. Define $\tilde{m}_d^{\tilde{\epsilon}, \Sigma}$ the same as $\tilde{m}_d^{\Sigma}$ in (5) by using $\tilde{\epsilon}_{ij} \equiv Y_{ij} - m^{\text{add}}(X_{i1j}, \dots, X_{iDj}) = m(X_{i1j}, \dots, X_{iDj}) - m^{\text{add}}(X_{i1j}, \dots, X_{iDj}) + \epsilon_{ij}$ instead of $Y_{ij} - \hat{m}_0$.

Proposition 1—Under assumptions A1–A5 with $m_{d,\text{true}}$ replaced by m_d^* , we have, for $d = 1, \dots, D$, for the local linear estimator,

$$\sup_{x_d \in [0,1]} \left| \widehat{m}_d(x_d) - m_d^*(x_d) - \beta_d^{LL}(x_d) - \tilde{m}_d^{\tilde{\epsilon}, \Sigma}(x_d) \right| = o_p(n^{-2/5}).$$

Remark 3: It is easy to see that

$$n^{2/5} \tilde{m}_d^{\tilde{\epsilon}, \Sigma}(x_d) \xrightarrow{d} \text{Normal} \left\{ 0, c_d^{-1} \mu_0 \{ (K^0)^2 \} \text{trace} \{ \Sigma^{-1} (\Sigma_{\text{true}} + \sum_d^m) \Sigma^{-1} P^d(x_d) \} \left\{ \sum_{j=1}^J b_{jj, \Sigma} p_j^d(x_d) \right\}^{-2} \right\}$$

where $\sum_d^m = \text{cov}_d^* \{ (m(X_{i11}, \dots, X_{iDj}) - m^{\text{add}}(X_{i11}, \dots, X_{iDj}))_{j=1, \dots, J}^{\top} \}$ and where cov_d^* means the conditional covariance matrix given $X_{1d1} = x_d, \dots, X_{1dJ} = x_d$. The expansions coincide with the case when the true regression function is additive. However, when it is not additive, in the variance expression the covariance Σ_{true} is replaced by $\Sigma_{\text{true}} + \Sigma_d^m$. This additional price for

deviations from additivity also appeared in applications of smooth backfitting to spatial data (Lu et al. 2007).

Remark 4: It is important to remember that when the true regression function is not additive, an additive regression model does not consistently estimate the true regression function, but estimates $m_0 + \sum_{d=1}^D m_d^*(x_d)$, which is the projection of the true regression function onto the space of additive models.

3 Profiling Estimator in Partially Linear Additive Models

3.1 The Method

In many studies, it is useful and interesting to impose a linear relation between the response and some covariates as in (2). As we now show, estimation of θ_{true} is easily accomplished by a profiling argument.

For any random sample $(W_1, \dots, W_n)$, let $\widehat{m}_0^W$ and $\{\widehat{m}_d^W(\cdot)\}_d$ be the (componentwise) smooth backfitting estimate when treating W as the responses and let $m_d^{*,W}(\cdot)$ be the limit of the latter. Here we suppress the dependence of the estimation on Σ . Define

$\widehat{U}_{i\ell j} = Z_{i\ell j} - \widehat{m}_0^{Z_\ell} - \sum_{d=1}^D \widehat{m}_d^{Z_\ell}(X_{idj})$, $\widehat{\mathcal{U}}_i = (\widehat{U}_{i\ell j})_{\ell=1, \dots, p; j=1, \dots, J}$ and $\widehat{S}_Z(\Sigma) = n^{-1} \sum_{i=1}^n \widehat{\mathcal{U}}_i \Sigma^{-1} \widehat{\mathcal{U}}_i^\top$. Mimicking Lin and Carroll (2006), for any working covariance matrix Σ , the profile estimator is easily found to have the closed form expression

$$\widehat{\theta} = \widehat{S}_Z^{-1}(\Sigma) n^{-1} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk, \Sigma} \widehat{U}_{ik} \{Y_{ij} - \widehat{m}_0^Y - \sum_{d=1}^D \widehat{m}_d^Y(X_{idj})\}$$

where $\widehat{U}_{ik} = (\widehat{U}_{i1k}, \dots, \widehat{U}_{ipk})^\top$.

Remark 5—Note crucially here that $\widehat{U}_{ik}$ is the result of an additive regression of Z in X , even though that regression need not be additive in actuality. This is where the result in Section 2.3 comes into play.

3.2 Asymptotic Theory

Define

$U_{i\ell j} = Z_{i\ell j} - m_0^{*,Z_\ell} - \sum_{d=1}^D m_d^{*,Z_\ell}(X_{idj})$, $\mathcal{U}_i = (U_{i\ell j})_{\ell=1, \dots, p; j=1, \dots, J}$, $U_{ik} = (U_{i1k}, \dots, U_{ipk})^\top$ and $S_Z(\Sigma) = E(\mathcal{U}_i \Sigma^{-1} \mathcal{U}_i^\top)$. Then, by using Proposition 1 and orthogonality properties in the calculations, we have the following result, sketched in Appendix A.2.

Proposition 2—Suppose $S_Z(\Sigma)$ is invertible. Make the assumptions of Proposition 1 for the additive regressions with covariates $\mathbf{X}_i$ and with responses $\mathbf{Y}_i$ and $Z_{i\ell} = (Z_{i\ell 1}, \dots, Z_{i\ell j})$ for $\ell = 1, \dots, p$. We also assume that the kernel function K^0 has bounded derivatives and $E\{\exp(|Z_{i\ell j}|)\} < \infty$. Then we have

$$\begin{aligned} n^{1/2}(\widehat{\theta} - \theta_{\text{true}}) &= S_Z^{-1}(\Sigma) n^{-1/2} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk, \Sigma} U_{ik} \epsilon_{ij} + o_p(1) \\ &= S_Z^{-1}(\Sigma) n^{-1/2} \sum_{i=1}^n \mathcal{U}_i \Sigma^{-1} \tilde{\epsilon}_i + o_p(1) \\ &\Rightarrow \text{Normal}\{0, S_Z^{-1}(\Sigma) E(\mathcal{U} \Sigma^{-1} \Sigma_{\text{true}} \Sigma^{-1} \mathcal{U}^\top) S_Z^{-1}(\Sigma)\}. \end{aligned} \quad (8)$$

When the working covariance matrix equals the true covariance matrix, i.e., $\Sigma = \Sigma_{\text{true}}$,

$$n^{1/2}(\hat{\theta} - \theta_{\text{true}}) \Rightarrow \text{Normal}\{0, S_z^{-1}(\Sigma_{\text{true}})\}. \quad (9)$$

Remark 6: Plug-in estimators of the covariance matrices in (8) and (9) can be constructed easily. For instance, $\mathcal{U}$ and $S_Z(\Sigma_{\text{true}})$ can be estimated by $\hat{\mathcal{U}}$ and $\hat{S}_Z(\Sigma_{\text{true}})$ as in Section 3.1, thus giving an estimator for the covariance matrix in (9). An estimator of Σ_{true} can be constructed using the sample covariance of residuals from a nonparametric additive regression using working independence. For the covariance matrix in (8), the term $E(\mathcal{U}\Sigma^{-1}\Sigma_{\text{true}}\Sigma^{-1}\mathcal{U}^\top)$ can be estimated by the sample covariance matrix of the estimates of $\mathcal{U}_i\Sigma^{-1}\epsilon_i$.

3.3 Semiparametric Efficiency

For Gaussian errors the semiparametric efficiency bound for the variance can be shown to equal $S_z^{-1}(\Sigma_{\text{true}})$. Thus in this case $\hat{\theta}$ is asymptotically efficient. Here we state the main details.

Consider the model given in (2) where $\tilde{\epsilon}_i = \text{Normal}(0, \Sigma_{\text{true}})$. Recall that $\Sigma_{\text{true}}^{-1} = (b_{jk, \text{true}})_{j,k=1, \dots, J}$. Define the symmetric matrix $\Sigma_{\text{true}}^{-1/2} = (a_{jk, \text{true}})_{j,k=1, \dots, J}$.

Proposition 3—Assume that $\tilde{\epsilon} = \text{Normal}(0, \Sigma_{\text{true}})$. Given $\Sigma = \Sigma_{\text{true}}$, the semiparametric efficient influence function for θ_0 is given by

$$\Psi_{\theta, \text{eff}} = S_z^{-1}(\Sigma_{\text{true}}) \sum_{j=1}^J \sum_{k=1}^J b_{jk, \text{true}} U_k \epsilon_j = S_z^{-1}(\Sigma_{\text{true}}) \mathcal{U} \Sigma_{\text{true}}^{-1} \tilde{\epsilon}.$$

The semiparametric variance bound for θ_0 is given by $E(\Psi_{\theta, \text{eff}} \Psi_{\theta, \text{eff}}^\top) = S_z^{-1}(\Sigma_{\text{true}})$, and in this case our estimator is semiparametric efficient.

4 Simulations

In this section, we discuss finite sample properties of the proposed estimators via simulation studies. Compared to the ordinary partially linear model, the models considered in this paper have two important characteristics: one is that we observe independent clusters which may have within-cluster correlations and the other is that the functional parameter in the regression model has an additive structure. Our method considers these two aspects of the model. We will compare the finite sample performance of our estimator and three alternatives, as follows.

- A. Our proposed method which considers within-cluster correlations and the additive structure of the nuisance parameter
- B. An estimator which consider only the additive structure of the nuisance parameter, but assumes that there is no correlation within the clusters. This is the working independence estimator for an additive model (Yu et al 2009).
- C. An estimator which consider only within-cluster correlations, but does not model the nonparametric function additively. This method is based on the estimators of Wang (2003) and Wang, et al. (2005), but instead uses a multivariate kernel function.
- D. An estimator which consider neither the correlations nor the additive structure, i.e., the ordinary profile multivariate-kernel estimator.

We considered the semiparametric additive regression model

$$Y_{ij} = \beta Z_{ij} + m_1(X_{i1j}) + m_2(X_{i2j}) + \epsilon_{ij}$$

where $\beta = 1.5$, $m_1(x) = \sin\{2\pi(x - 0.5)\}$ and $m_2(x) = x - 0.5 + \sin\{2\pi(x - 0.5)\}$. The sample size was 200, the bandwidth was 0.1, there were $J = 3$ repeated measures and the grid size for integration was 0.01. For each scenario we generated 500 data sets. Estimation of the covariance matrix used the residuals from the pooled data estimator. We generated $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3})$ from $\text{Normal}(0, \Sigma_E)$ where Σ_E has elements σ_{ij} . We investigated seven cases. For Cases 1, 2 and 3, we used the exchangeable covariance matrix $\Sigma_E = (1 - \rho_a)I + \rho_a \mathbf{1}\mathbf{1}^T$ with $\rho_1 = 0.9$, $\rho_2 = 0.5$ and $\rho_3 = 0.1$, respectively. For Case 4, we used common variances $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$ with $\sigma_{12} = 0.9$, $\sigma_{13} = 0.5$ and $\sigma_{23} = 0.4$. For Case 5, we used common variances $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$ with AR(1) structure having correlation coefficient -0.9 . Finally, for Cases 6 and 7 we allowed heteroscedasticity with $\sigma_{11} = 9$, $\sigma_{22} = 4$ and $\sigma_{33} = 1$, and with common correlation 0.9 and 0.1, respectively.

The six dimensional vector X_i was generated from independent $\text{Normal}(0.5, 0.25)$ but truncated to the 6 dimensional unit cube $[0, 1]^6$. The covariate Z_{ij} was generated as $Z_{ij} = 3(1 - 2X_{i1j})(1 - 2X_{i2j}) + u_{ij}$ where u_{ij} were independently from $\text{Normal}\{0, 0.5\}$. Note that the function $f(x, y) = (1 - 2x)(1 - 2y)$ is orthogonal to additive functions under this setup so that

$E\{f(X_{i11}, X_{i21}), \dots, f(X_{i1J}, X_{i2J})\} \Sigma_{\text{true}}^{-1} \{f(X_{i11}, X_{i21}), \dots, f(X_{i1J}, X_{i2J})\}^T$ is the source of our gaining efficiency by considering additivity. To see this, recall that the asymptotic variance of type C estimator is $E(\mathcal{V}_i \Sigma_{\text{true}}^{-1} \mathcal{V}_i^T)^{-1}$ where $\mathcal{V}_i = \{Z_{i1} - E(Z_{i1}|X_{i11}, X_{i21}), \dots, Z_{iJ} - E(Z_{iJ}|X_{i1j}, X_{i2j})\}$. Then

$$\begin{aligned} & E(\mathcal{U}_i \Sigma_{\text{true}}^{-1} \mathcal{U}_i^T) \\ &= E(\mathcal{V}_i \Sigma_{\text{true}}^{-1} \mathcal{V}_i^T) \\ & \quad + E(f(X_{i11}, X_{i21}), \dots, f(X_{i1J}, X_{i2J})) \Sigma_{\text{true}}^{-1} (f(X_{i11}, X_{i21}), \dots, f(X_{i1J}, X_{i2J}))^T \end{aligned}$$

because the function $f(x, y) = E(Z|X_1 = x, X_2 = y)$ is orthogonal to additive functions.

We report the mean squared errors (MSE) for estimating β in Table 1. The biases were negligible in almost every case and are not reported here. The worst bias was about -0.015 for the local linear estimator of type D in Case 7. Here is a summary of the main findings.

- There is little difference between the local linear and local constant estimators.
- Our method that takes into account both the additivity and the within-cluster correlation does much better than the other methods.
- The comparison of whether to ignore the additivity or the within-cluster correlation, methods B and C, depends on the covariance structure.
- The method that ignores both the additivity and the within-cluster correlations is wildly inefficient.

5 Urinary Nitrogen Example

To illustrate the repeated measures smooth backfitting algorithm, we use data from the OPEN Study (Kipnis et al 2003). Background on nutritional epidemiology may be found in Willett (1990). The data was analyzed in Carroll et al (2009) with the additive model framework but no parametric part.

We use a data set of 294 men and women measured at two visits who reported their short-term intake of protein Y_{ij} as measured by the biomarker urinary nitrogen. Here the protein biomarker

data were log-transformed. To predict protein intake we used three variables, body mass index X_{i1j} , log-protein intake X_{i2j} as measured by a 24-hour recall instrument and log-protein intake Z_{ij} as measured by a food frequency questionnaire. A preliminary analysis using three dimensional nonparametric additive fit suggested linearity in the last predictor Z_{ij} . The residuals from an additive regression fit suggested an estimated covariance matrix with variances 0.065 and 0.075 and with a correlation of 0.507. The bandwidths were selected as in Carroll et al (2009) and are given as 3:02 for body mass index X_{i1j} and 0.374 for log-protein via the 24-hour recall X_{i2j} .

Under this working covariance, we fitted the following model:

$$E(Y_{ij}|Z_{ij}, X_{i1j}, X_{i2j}) = m_0 + \beta Z_{ij} + m_1(X_{i1j}) + m_2(X_{i2j}).$$

We also found that third order and second order polynomials approximate $m_1(\cdot)$ and $m_2(\cdot)$ fairly well, respectively. Table 2 shows the estimates of β from the parametric polynomial fits and semiparametric fits with (WD) and without (WI) taking account the within-cluster correlations. As is seen in the table, they give similar results with slightly smaller standard errors when taking account of the within-cluster correlations. We also found that the average length of pointwise confidence interval of WD estimators for nonparametric functions were 13.5% shorter than WI estimators. Figure 1 shows the residual plots and Figure 2 shows the fits of the functions $m_1(\cdot)$ and $m_2(\cdot)$.

6 Score Testing for Parametric Effects in a Complex Interaction Model

Although our major focus is on the partially linear repeated measures additive model, the power of our results and another application of Proposition 1 can be seen in their application to a unique testing problem not considered previously in the additive models case. In a genetics context (Chatterjee et al. 2006), there is a parametric modeling framework for testing certain parametric components in a way that gains considerable power in the presence of interactions. The natural version of their model in our framework is as follows. We have additional covariates Z_{ij} , and the model consists of a parametric component in the Z_{ij} , an additive model in the X_{idj} and a possible interaction determined by a scalar γ , and is

$$Y_{ij} = m_{0,\text{true}} + \sum_{d=1}^D m_{d,\text{true}}(X_{idj}) + \{1 + \gamma \sum_{d=1}^D m_{d,\text{true}}(X_{idj})\} Z_{ij}^\top \theta_{\text{true}} + \epsilon_{ij}. \quad (10)$$

Here the goal is to test whether Z influences Y , and thus the null hypothesis is that $H_0: \theta_{\text{true}} = 0$, see Chatterjee et al. for detailed motivation. Basically, the idea is that in some cases it can be expected that the Z 's and the X 's interact, and model (10) is one way to exploit this expectation to obtain a more powerful test for the main effect of Z .

The value of the interaction γ is not identified under the null hypothesis, and Chatterjee et al. refer to this as a Tukey-type formulation, with the exception that the interest is in the main effect and not in the interaction.

Because (10) is a complex model that is not even identified under the null model, Chatterjee et al. suggest using a modified score test procedure. In our case, even more is true, because even if we fix $\gamma \neq 0$, (10) is a non-standard model for which there are no available methods. Maity, et al. (2009) consider this problem when there are no repeated measures, and our development is a generalization of theirs.

The normalized loglikelihood score statistic in θ for model (10) evaluated at $\theta_{\text{true}} = 0$ is easily seen to equal

$$\mathcal{S}_n(\gamma) = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^J b_{jk, \Sigma_{\text{true}}} \{1 + \gamma \sum_{d=1}^D \widehat{m}_d(X_{ijd})\} Z_{ij} \{Y_{ik} - \widehat{m}_0 - \sum_{s=1}^D \widehat{m}_s(X_{isk})\},$$

where $\{\widehat{m}_d(\cdot)\}_d$ are computed at the null model with $\theta_{\text{true}} = 0$. Chatterjee et al. (2006) suggest normalizing the score statistic $\mathcal{S}_n(\gamma)$ so that it is asymptotically chi-square, and then maximizing over an interval of γ -values.

One can derive uniform expansions of the function estimates similar to that of Lin and Carroll (2006) to order $o_p(n^{-1/2})$ and use those to derive the asymptotic distribution of $\mathcal{S}_n(\gamma)$ under the null hypothesis that $\theta_{\text{true}} = 0$. However, two difficulties arise. The first is that undersmoothing is required, i.e., $nh_d^4 \rightarrow 0$ is needed for $d = 1, \dots, D$. Second, and critically, the distribution depends upon terms that are not directly computable because they are the solution to an integral equation, and hence computing p-values in simple and easily computed ways is not possible.

6.1 An Adjusted Score

To overcome both problems raised above, we use the following projection-type approach.

Define $Q_{ij}(\gamma) = \{1 + \gamma \sum_{d=1}^D m_{d, \text{true}}(X_{ijd})\} Z_{ij}$ and $\widehat{Q}_{ij}(\gamma) = \{1 + \gamma \sum_{d=1}^D \widehat{m}_d(X_{ijd})\} Z_{ij}$, where again $\{\widehat{m}_d(\cdot)\}_d$ are computed in model (10) with $\theta_{\text{true}} = 0$.

Let $\widehat{m}_d^Q(\cdot, \gamma)$ and $\check{m}_d^Q(\cdot, \gamma)$ be the smooth backfitting estimators when the "responses" are $Q_{ij}(\gamma)$ and $\widehat{Q}_{ij}(\gamma)$, respectively, and similarly for $\widehat{m}_0^Q$ and $\check{m}_0^Q$. We emphasize that we are not assuming that $Q_{ij}(\gamma)$ follows an additive model, and *this is where Proposition 1 comes into play*.

We propose to use the adjusted score

$$\mathcal{S}_{n, \text{adj}}(\gamma) = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^J b_{jk, \Sigma_{\text{true}}} \{\widehat{Q}_{ij}(\gamma) - \check{m}_0^Q - \sum_{d=1}^D \check{m}_d^Q(X_{ijd}, \gamma)\} \times \{Y_{ik} - \widehat{m}_0 - \sum_{s=1}^D \widehat{m}_s(X_{isk})\}.$$

Define $U_{ijQ}(\gamma) = Q_{ij} - m_0^Q - \sum_{d=1}^D m_d^Q(X_{ijd}, \gamma)$ and $\mathcal{U}_{iQ}(\gamma) = \{U_{i1Q}(\gamma), \dots, U_{iJQ}(\gamma)\}$. Then we have the following result, which is the first step in the construction of the test-statistic, namely normalization to a chi-squared random variable for fixed γ .

Proposition 4—*Suppose that the dimension of Z is p_w . Assume that the assumptions of Proposition 1 holds for the additive regressions with covariates $\mathbf{X}_i$ and responses $\mathbf{Y}_i$ and $(Q_{i1}, \dots, Q_{iJ})$. We also assume that the kernel function K^0 has bounded derivative and $E \exp(|Q_{ij}|) < \infty$. Then, under the null hypothesis, we have*

$$\begin{aligned} \mathcal{S}_{n, \text{adj}}(\gamma) &= n^{-1/2} \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^J b_{jk, \Sigma_{\text{true}}} U_{ijQ}(\gamma) \epsilon_{ik} + o_p(1) \\ &\rightarrow \text{Normal}[0, E\{\mathcal{U}_Q(\gamma) \Sigma_{\text{true}}^{-1} \mathcal{U}_Q^{\top}(\gamma)\}]. \end{aligned} \quad (11)$$

Let $\widehat{\Sigma}_{\text{true}}^{-1}$ with elements $\widehat{b}_{jk}$ be a consistent estimate of $\Sigma_{\text{true}}^{-1}$ and define $\widehat{U}_{ijQ}(\gamma) = \widehat{Q}_{ij} - \widehat{m}_0^Q - \sum_{d=1}^D \widehat{m}_d^Q(X_{idj}, \gamma)$. Then

$$\widehat{\Lambda}(\gamma) = n^{-1} \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^J \widehat{b}_{jk} \widehat{U}_{ijQ}(\gamma) \widehat{U}_{ikQ}^T(\gamma)$$

is a consistent estimate of $E\{\mathcal{U}_Q(\gamma) \Sigma_{\text{true}}^{-1} \mathcal{U}_Q^T(\gamma)\}$, and thus $\mathcal{T}_n(\gamma) = \mathcal{S}_{n,\text{adj}}^T(\gamma) \widehat{\Lambda}^{-1}(\gamma) \mathcal{S}_{n,\text{adj}}(\gamma)$ is asymptotically chisquared with p_w degrees of freedom.

6.2 Implementation and p-Values

Chatterjee et al. suggest the following procedure. Instead of choosing an arbitrary γ , they form $n = \max_{0 \leq \gamma \leq b} n(\gamma)$. We emphasize that they do not consider our model, only parametric ones. It is possible to show that $n(\gamma)$ converges weakly to a Gaussian process in γ . Hence, at least in principle, it is possible to use this fact directly to obtain p-values, but the process is not easy because it requires sampling from the Gaussian process.

An alternative is to use the parametric bootstrap, by assuming that the (ϵ_{ij}) are normally distributed and generating bootstrap samples under the null hypothesis for model (10) evaluated at $\theta_0 = 0$, with the fits $\{\widehat{m}_d(\cdot)\}$ replacing the true functions. While simple, this relies on the normality assumption and requires computing smoothing backfitting estimates for all the bootstrap samples.

To avoid these difficulties, we reply on the linear expansion (11). Then we propose to use n as the test statistic. Its percentiles can be estimated as follows. Let B be a large number. For $b = 1, \dots, B$, generate mean zero normal random variables $\mathcal{Z}_{bij}$ such that $(b_{i1}, \dots, b_{iJ})^T = \text{Normal}(0, \Sigma_{\text{true}})$. Then compute

$$\begin{aligned} \mathcal{S}_{bn,\text{adj}}(\gamma) &= n^{-1/2} \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^J b_{jk, \Sigma_{\text{true}}} \widehat{U}_{ijQ}(\gamma) z_{bik}; \\ \mathcal{T}_{bn} &= \max_{0 \leq \gamma \leq b} \mathcal{S}_{bn,\text{adj}}^T(\gamma) \widehat{\Lambda}^{-1}(\gamma) \mathcal{S}_{bn,\text{adj}}(\gamma). \end{aligned} \quad (12)$$

Note that except for the estimation of U_{ijQ} , the limiting distribution of $\mathcal{S}_{bn,\text{adj}}(\gamma)$ is the same as the limiting distribution of $\mathcal{S}_{n,\text{adj}}(\gamma)$, no matter what the distribution of the (ϵ_{ij}) is. This

means that a consistently estimated p-value for the null hypothesis is $B^{-1} \sum_{b=1}^B I(\mathcal{T}_n < \mathcal{T}_{bn})$.

Remark—It is worth emphasizing that in constructing $\mathcal{S}_{bn,\text{adj}}(\gamma)$ for $b = 1, \dots, B$ in (12), smooth backfitting is not used. Our methods require only that smooth backfitting for repeated measures be applied to the actual data, only once, to construct $\widehat{m}_d(\cdot)$ and $\widehat{m}_d^Q(\cdot, \gamma)$. Thus, computing the p-value is not a difficult process.

7 Discussion and Conclusions

We have discussed fitting semiparametric additive models in the repeated measures context. We have used the smooth backfitting kernel approach to this problem largely because it has a tractable asymptotic theory, with good efficiency properties. In practice, of course, regression spline approaches can be used as well. In simulations not reported here, we have found that regression spline approaches have similar good small-sample behavior, although asymptotic theory would be challenging.

Acknowledgments

Yu and Mammen's research was supported by the Deutsche Forschungsgemeinschaft project MA 1026/7-3. Carroll and Maity's research was supported by a grant from the National Cancer Institute (CA57030). Carroll's work was also supported by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST).

Appendix: Proofs and Complements

A.1 Proof of Proposition 1

Let $\mathcal{F} = \{\mathbf{f} = (f_1, \dots, f_J)^\top : f_j \text{ are real valued functions on } [0, 1]^D\}$ and

$\mathcal{H} = \{\mathbf{f} = (f_1, \dots, f_J)^\top : f(x) = \sum_{d=1}^D f_d(x_d) \text{ is additive functions on } [0, 1]^D\}$. Note that $\mathcal{H}$ is a subspace of $\mathcal{F}$ and has one common function f . We equip $\mathcal{F}$ with a Hilbert (semi-)norm $\|\mathbf{f}\|_\Sigma = \int \mathbf{f}^\top(\mathbf{x}) \Sigma^{-1} \mathbf{f}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$. Let also $\mathcal{H}^0 = \{\mathbf{f} \in \mathcal{H} : \langle \mathbf{1}, \mathbf{f} \rangle_\Sigma = 0\}$ where $\mathbf{1} = (1, \dots, 1)^\top \in \mathcal{F}$ and $\langle \cdot, \cdot \rangle_\Sigma$ is the inner product inducing the norm $\|\cdot\|_\Sigma$. Let $\check{m}(\mathbf{x}) = \{m(\mathbf{x}_1), \dots, m(\mathbf{x}_J)\}^\top \in \mathcal{F}$ with $\mathbf{x}_j = (x_{1j}, \dots, x_{Dj})^\top$. For simplicity, we assume that the $\langle \check{m}, \mathbf{1} \rangle_\Sigma = 0$, i.e., it is centered.

Define $m^*(\mathbf{x}) = \{\sum_{d=1}^D m_d^*(x_{d1}), \dots, \sum_{d=1}^D m_d^*(x_{dJ})\}^\top$ as the projection of $\check{m}$ onto $\mathcal{H}_0$, i.e. $m^* = \arg\min_{\mathbf{h} \in \mathcal{H}_0} \|\check{m} - \mathbf{h}\|_\Sigma$. This yields that

$$\begin{aligned} 0 &= \langle \check{m} - m^*, \mathbf{g} \rangle_\Sigma \\ &= \int \{m(\mathbf{x}_1) - \sum_{d=1}^D m_d^*(x_{d1}), \dots, m(\mathbf{x}_J) - \sum_{d=1}^D m_d^*(x_{dJ})\} \\ &\quad \times \Sigma^{-1} \{\sum_{d=1}^D g_d(x_{d1}), \dots, \sum_{d=1}^D g_d(x_{dJ})\}^\top p(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (\text{A.1})$$

for any $\mathbf{g}(\mathbf{x}) = (\sum_{d=1}^D g_d(x_{d1}), \dots, \sum_{d=1}^D g_d(x_{dJ}))^\top \in \mathcal{H}_0$ for which the integral exists. We can derive following system of integral equations from the equation (A.1):

$$\begin{aligned} m_d^*(x_d) &= m_d^\Sigma(x_d) - \int m_d^*(t) \frac{\sum_{j=1}^J \sum_{k \neq \ell} b_{jk, \Sigma} p_{jk}^{dd}(x_d, t)}{\sum_{j=1}^J b_{jj, \Sigma} p_j^d(x_d)} dt \\ &\quad - \sum_{s=1, s \neq d}^D \int m_s^*(t) \frac{\sum_{j=1}^J \sum_{k=1}^J b_{jk, \Sigma} p_{jk}^{ds}(x_d, t)}{\sum_{j=1}^J b_{jj, \Sigma} p_j^d(x_d)} dt \end{aligned} \quad (\text{A.2})$$

$$\text{where } m_d^\Sigma(x_d) = \sum_{j=1}^J \sum_{k=1}^J b_{jk, \Sigma} E(Y_{ik} | X_{idj} = x_d) p_j^d(x_d) / \sum_{j=1}^J b_{jj, \Sigma} p_j^d(x_d).$$

The proof of Proposition 1 can be done as in Carroll et al. (2009) based on the orthogonality that arises in (A.1) and modified definition of functions they call $\tilde{m}_d^{V, M}(x_d)$ and $\tilde{m}_d^{V, E}(x_d)$, which are given as

$$\begin{aligned} \tilde{m}_d^{V, M}(x_d) &= n^{-1} \sum_{i=1}^n \sum_{j,k=1}^J b_{jk, V} \sum_{s=1}^D m_s^*(X_{isk}) K_{h_d}(x_d, X_{idj}) / \sum_{j=1}^J b_{jj, V} \tilde{p}_j^d(x_d); \\ \tilde{m}_d^{V, E}(x_d) &= n^{-1} \sum_{i=1}^n \sum_{j,k=1}^J b_{jk, V} \check{\epsilon}_{ik} K_{h_d}(x_d, X_{idj}) / \sum_{j=1}^J b_{jj, V} \tilde{p}_j^d(x_d). \end{aligned}$$

Note that the orthogonality in equation (A.1) implies that

$$E\{\sum_{j,k=1}^J b_{jk,\Sigma} \check{\epsilon}_{1k} K_{h_d}(x_d, X_{1dj})\} = 0 \text{ for } d=1, \dots, D.$$

A.2 Sketch Proof of Proposition 2 and 4

The basic idea of the proof is related to that of Yu et al. (2009). By Proposition 1, we have

$$\widehat{S}_z(\Sigma) \xrightarrow{P} S_z(\Sigma). \text{ It suffices to show that}$$

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk,\Sigma} \widehat{U}_{ik} \{Y_{ij} - \widehat{m}_0^Y - \sum_{d=1}^D \widehat{m}_d^Y(X_{idj}) - \widehat{U}_{ij}^\top \theta_{\text{true}}\} \\ &= n^{-1/2} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk,\Sigma} U_{ik} \epsilon_{ij} + o_p(1). \end{aligned}$$

Note that

$$Y_{ij} - \widehat{m}_0^Y - \sum_{d=1}^D \widehat{m}_d^Y(X_{idj}) - \widehat{U}_{ij}^\top \theta_{\text{true}} = m_{0,\text{true}} - \widehat{m}_0^{Y(\beta_{\text{true}})} + \sum_{d=1}^D m_{d,\text{true}}(X_{idj}) - \sum_{d=1}^D \widehat{m}_d^{Y(\beta_{\text{true}})}(X_{idj}) + \epsilon_{ij} \text{ where } Y(\beta_{\text{true}}) = Y - Z^\top \beta_{\text{true}} (=Y - Z^\top \beta)$$

since the estimators are additive in responses. Now, we have

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk,\Sigma} \widehat{U}_{ik} \{Y_{ij} - \widehat{m}_0^Y - \sum_{d=1}^D \widehat{m}_d^Y(X_{idj}) - \widehat{U}_{ij}^\top \theta_{\text{true}}\} \\ &= n^{-1/2} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk,\Sigma} U_{ik} \epsilon_{ij} \\ & \quad + n^{-1/2} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk,\Sigma} U_{ik} \{m_{0,\text{true}} - \widehat{m}_0^{Y(\beta_{\text{true}})} + \sum_{d=1}^D (m_{d,\text{true}}(X_{idj}) - \widehat{m}_d^{Y(\beta_{\text{true}})}(X_{idj}))\} \\ & \quad + n^{-1/2} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk,\Sigma} \{\widehat{U}_{ik} - U_{ik}\} \epsilon_{ij} \\ & \quad + n^{-1/2} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk,\Sigma} \{\widehat{U}_{ik} - U_{ik}\} \\ & \quad \quad \times \{m_{0,\text{true}} - \widehat{m}_0^{Y(\beta_{\text{true}})} + \sum_{d=1}^D (m_{d,\text{true}}(X_{idj}) - \widehat{m}_d^{Y(\beta_{\text{true}})}(X_{idj}))\}. \\ & \equiv n^{-1/2} \sum_{i=1}^n \sum_{k=1}^J \sum_{j=1}^J b_{jk,\Sigma} U_{ik} \epsilon_{ij} + C_{n1} + C_{n2} + C_{n3}. \end{aligned}$$

It is clear that C_{n2} and C_{n3} are $o_p(1)$ by Proposition 1. Now we will show $C_{n1} = o_p(1)$. By Proposition 1, we have

$$\sup_{x \in [0,1]^D} \left| m_{0,\text{true}} - \widehat{m}_0^{Y(\beta_{\text{true}})} + \sum_{d=1}^D \{m_{d,\text{true}}(x_d) - \widehat{m}_d^{Y(\beta_{\text{true}})}(x_d)\} \right| = o_p(\delta_n) \quad (\text{A.3})$$

for $\delta_n = n^{-1/5+a}$ with any small positive a . Denote

$$\widehat{h}_n(x_1, \dots, x_d) = m_{0,\text{true}} - \widehat{m}_0^{Y(\beta_{\text{true}})} + \sum_{d=1}^D (m_{d,\text{true}}(x_d) - \widehat{m}_d^{Y(\beta_{\text{true}})}(x_d)) \equiv \widehat{h}_{n,0} + \sum_{d=1}^D \widehat{h}_{n,d}(x_d).$$

Using definition of $\widehat{m}_d(x_d)$, $d = 1, \dots, D$ and results from Carroll et al (2009), we have

$$\begin{aligned} \frac{d}{dx_d} \widehat{m}_d(x_d) &= \frac{d}{dx_d} \widehat{m}_d^\Sigma(x_d) - \int \widehat{m}_d(t) \frac{d}{dx_d} \frac{\sum_{j=1}^J \sum_{k \neq \ell}^J b_{jk,\Sigma} \widehat{p}_{jk}^{dd}(x_d, t)}{\sum_{j=1}^J b_{jj,\Sigma} \widehat{p}_j^d(x_d)} dt \\ & \quad - \sum_{s=1, s \neq d}^D \int \widehat{m}_s(t) \frac{d}{dx_d} \frac{\sum_{j=1}^J \sum_{k=1}^J b_{jk,\Sigma} \widehat{p}_{jk}^{ds}(x_d, t)}{\sum_{j=1}^J b_{jj,\Sigma} \widehat{p}_j^d(x_d)} dt \end{aligned} \quad (\text{A.4})$$

if $\sum_{j=1}^J b_{jj,\Sigma} \widehat{p}_j^d(x_d) > 0$, which happens with probability tending to one. By Proposition 1, standard calculations in kernel smoothing and integration by part, we have

$$\sup_{x_d \in [0,1]} \left| \frac{d}{dx_d} \widehat{h}_{n,d}(x) - h_d b_{n,d}(x_d) \right| = O_p(n^{-3/10} \sqrt{\log n}) = o_p(\delta_n) \quad (\text{A.5})$$

for some uniformly bounded non-random functions $b_{n,d}$. Equation (A.3) and (A.5) imply that

$\delta_n^{-1} \widehat{h}_n(\cdot) \in B(\mathbf{0}, 1) \equiv \{g = \sum_{d=1}^D g_d : g_d \text{ are real functions on } [0, 1] \text{ with } \|g\|_\infty \leq 1 \text{ and } \sup_{x_d, y_d \in [0,1]} |g_d(x_d) - g_d(y_d)| \leq |x_d - y_d| \text{ for } d = 1, \dots, D\}$ with probability tending to one. Note that the covering number with bracketing $N(\eta) \equiv N\{\eta, B(\mathbf{0}, 1), \|\cdot\|_\infty\}$ of $B(\mathbf{0}, 1)$ with respect to the uniform norm is bounded by $(2\eta^{-1})^{D3D\eta^{-1}}$, i.e., one can find a set $\mathcal{N}(\eta)$ that has $N(\eta)$ elements such that for any $g \in B(\mathbf{0}, 1)$ there exists $g_i^L, g_i^U \in \mathcal{N}(\eta)$ with $g_i^L \leq g \leq g_i^U$ and $\|g_i^L - g_i^U\|_\infty < \eta$.

Define $\{F(\mathbf{X}_i, \mathbf{Z}_{i\ell})\}_{g \in B} = \sum_{j=1}^J \sum_{k=1}^J b_{jk,\Sigma} U_{i\ell k} g(X_{i1j}, \dots, X_{iDj})$ for $g \in B$. Then

$F_n(\cdot) \equiv \left\{ n^{-1/2} \sum_{i=1}^n F(\mathbf{X}_i, \mathbf{Z}_{i\ell})(\cdot) \right\}$ can be regarded as a random process indexed by elements of B . Note that $E[\{F(\mathbf{X}_i, \mathbf{Z}_{i\ell})\}_{g \in B}] = 0$ for any $g \in B(\mathbf{0}, 1)$ by the orthogonality (A.1). Then one can show that $\sup_{g \in B(\mathbf{0}, 1)} |F_n g| = O_p(1)$. This follows by Corollary 8.8 in van de Geer (2000) with the above entropy condition and the tail condition assumed in the proposition.

Finally, we have $\Pr\{|\delta_n^{-1} C_{n1}| > M\} \leq \Pr\{\sup_{g \in B(\mathbf{0}, 1)} |F_n g| > M\} + \Pr\{\delta_n^{-1} \widehat{h}_n(\cdot) \notin B(\mathbf{0}, 1)\}$. Thus $C_{n1} = O_p(\delta_n) = o_p(1)$. This completes the proof. The same argument can be applied to the proof of Proposition 4.

A.3 Proof of Proposition 3

Recall that $\widehat{U}_{i\ell j} = Z_{i\ell j} - \widehat{m}_0^{Z_\ell} - \sum_{d=1}^D \widehat{m}_d^{Z_\ell}(X_{idj})$ and it is a residual of a nonparametric additive regression of $(Z_{i\ell})_{i=1, \dots, n}$ on $(X_{id})_{i=1, \dots, n}$. Hence using Carroll et al. (2009) we obtain that for any p ,

$$\begin{aligned} \sum_{\ell=1}^J \sum_{j=1}^J b_{\ell\ell,\text{true}} \widehat{p}_\ell^d(x_d) \widehat{m}_d^{Z_p}(x_d) &= n^{-1} \sum_{i=1}^n \sum_{\ell=1}^J \sum_{k=1}^J b_{\ell k,\text{true}} (Z_{ipk} - \widehat{m}_0) K_{h_d}(x_d, X_{id\ell}) \\ &\quad - \int \widehat{m}_d^{Z_p}(t) \sum_{\ell=1}^J \sum_{k \neq \ell=1}^J b_{\ell k,\text{true}} \widehat{p}_{\ell k}^{dd}(x_d, t) dt \\ &\quad - \int \sum_{a \neq d=1}^D \widehat{m}_a^{Z_p}(t) \sum_{\ell=1}^J \sum_{k=1}^J b_{\ell k,\text{true}} \widehat{p}_{\ell k}^{\text{da}}(x_d, t) dt. \end{aligned}$$

Now taking limits as $n \rightarrow \infty$ and simplifying, we obtain that for $d = 1, \dots, D$,

$$0 = \sum_{\ell=1}^J \sum_{k=1}^J b_{\ell k,\text{true}} p_\ell^d(x_d) E \left[Z_{pk} - m_0^{Z_p} - \sum_{a=1}^D m_a^{Z_p}(X_{ak}) \middle| X_{d\ell} = x_d \right]. \quad (\text{A.6})$$

A.3.1 Proof of Proposition 3: Semiparametric efficiency

Assume that $\tilde{\epsilon} \sim \text{Normal}(0, \Sigma_{\text{true}})$. Assuming that Σ_{true} is known, the loglikelihood (up to a constant) is given by

$$\begin{aligned}\mathcal{L} = & -(1/2) \sum_{j=1}^J \sum_{k=1}^J b_{jk,\text{true}} [Y_j - \{m_0 + Z_j^\top \theta_0 + \sum_{d=1}^D m_d(X_{dj})\}] \\ & \times [Y_k - \{m_0 + Z_k^\top \theta_0 + \sum_{d=1}^D m_d(X_{dk})\}].\end{aligned}$$

Parametric submodel, score functions and tangent space—Let M denote the true nonparametric model. Now we take a parametric submodel M_γ : for $\gamma = (\gamma_1, \dots, \gamma_D)$, we replace $m_d(X_{dk})$ by $m_d(X_{dk}, \gamma_d)$ and obtain the loglikelihood

$$\begin{aligned}\mathcal{L}_\gamma = & -(1/2) \sum_{j=1}^J \sum_{k=1}^J b_{jk,\text{true}} [Y_j - \{m_0 + Z_j^\top \theta_0 + \sum_{d=1}^D m_d(X_{dj}, \gamma_d)\}] \\ & \times [Y_k - \{m_0 + Z_k^\top \theta_0 + \sum_{d=1}^D m_d(X_{dk}, \gamma_d)\}].\end{aligned}$$

such that $\mathcal{L}_\gamma$ coincides with $\mathcal{L}$ for $\gamma = \gamma_0$. Under $\mathcal{L}_\gamma$, the score functions for m_0 , θ_0 and γ , evaluated at true values of θ_0 and m_0 and $\gamma = \gamma_0$, are given by

$$\begin{aligned}S_{m_0} &= \sum_{j=1}^J \sum_{\ell=1}^J b_{j\ell,\text{true}} \epsilon_j; \\ S_\theta &= \sum_{j=1}^J \sum_{\ell=1}^J Z_\ell b_{j\ell,\text{true}} \epsilon_j; \\ S_{\gamma_d} &= \sum_{j=1}^J \sum_{\ell=1}^J \eta_d(X_{d\ell}) b_{j\ell,\text{true}} \epsilon_j,\end{aligned}$$

where $\eta_d(X_{d\ell})$ are any function.

The Tangent Space—We see that the tangent space is spanned by S_{m_0} , S_θ and $(S_{\gamma_1}, \dots, S_{\gamma_D})$. We orthogonalize S_θ and the nuisance space $\mathcal{N}$ spanned by S_{m_0} and $(S_{\gamma_1}, \dots, S_{\gamma_D})$.

We start by observing that by (A.6) it is straightforward that for any function $(X_{d\ell})$,

$$0 = E \left[\sum_{\ell=1}^J \sum_{k=1}^J b_{\ell k,\text{true}} \mathcal{G}(X_{d\ell}) E \left\{ Z_{pk} - m_0^{Z_p} - \sum_{a=1}^D m_a^{Z_p}(X_{ak}) | X_{d\ell} \right\} \right]. \quad (\text{A.7})$$

Next we derive that for any d ,

$$\begin{aligned}E[S_\theta S_{\gamma_d}] &= E \left[\left\{ \sum_{j=1}^J \sum_{\ell=1}^J b_{j\ell,\text{true}} \epsilon_j Z_\ell \right\} \left\{ \sum_{k=1}^J \sum_{q=1}^J b_{kq,\text{true}} \epsilon_k \eta_d(X_{dq}) \right\} \right] \\ &= E \left[\sum_{j=1}^J \sum_{\ell=1}^J \sum_{k=1}^J \sum_{q=1}^J b_{j\ell,\text{true}} b_{kq,\text{true}} E \left\{ Z_\ell \eta_d(X_{dq}) \epsilon_j \epsilon_k | X_{dq} \right\} \right] \\ &= E \left\{ \sum_{\ell=1}^J \sum_{q=1}^J b_{\ell q,\text{true}} E(Z_\ell | X_{dq}) \eta_d(X_{dq}) \right\},\end{aligned}$$

the last line following from the fact that $\tilde{\epsilon} = \text{Normal}[0, \Sigma_{\text{true}}]$. Similarly we derive that

$$\begin{aligned} E \left[\left\{ \sum_{j=1}^J \sum_{\ell=1}^J b_{j\ell, \text{true}} \epsilon_j (m_0^{Z_p} + \sum_{a=1}^D m_a^{Z_p}(X_{a\ell})) \right\} S_{\gamma_d} \right] \\ = E \left[\sum_{\ell=1}^J \sum_{q=1}^J b_{\ell q, \text{true}} E \left\{ m_0^{Z_p} + \sum_{a=1}^D m_a^{Z_p}(X_{a\ell}) | X_{dq} \right\} \eta_d(X_{dq}) \right]. \end{aligned}$$

Hence, using (A.7) we see that for any d ,

$$\begin{aligned} E \left[\left\{ S_{\theta} - \sum_{j=1}^J \sum_{\ell=1}^J b_{j\ell, \text{true}} \epsilon_j (m_0^Z + \sum_{a=1}^D m_a^Z(X_{a\ell})) \right\} S_{\gamma_d} \right] \\ = E \left[\sum_{\ell=1}^J \sum_{q=1}^J b_{\ell q, \text{true}} \eta_d(X_{dq}) E \left\{ Z_{\ell} - m_0^Z - \sum_{a=1}^D m_a^Z(X_{a\ell}) | X_{dq} \right\} \right] = 0. \end{aligned}$$

Hence

$$S_{\theta}^{\Sigma_{\text{true}}} = \sum_{j=1}^J \sum_{\ell=1}^J b_{j\ell, \text{true}} \left\{ Z_{\ell} - m_0^Z - \sum_{a=1}^D m_a^Z(X_{a\ell}) \right\} \epsilon_j = \mathcal{U} \Sigma_{\text{true}}^{-1} \tilde{\epsilon}$$

is orthogonal to the nuisance space spanned by $(S_{\gamma_1}, \dots, S_{\gamma_D})$.

A similar argument can be used to prove that $S_{\theta}^{\Sigma_{\text{true}}}$ is also orthogonal to S_{m_0} .

Hence the efficient score for θ_0 is $S_{\theta}^{\Sigma_{\text{true}}}$, which in turn gives us the efficient influence function

$$\Psi_{\theta, \text{eff}} = S_z^{-1}(\Sigma_{\text{true}}) \mathcal{U} \Sigma_{\text{true}}^{-1} \tilde{\epsilon},$$

where $S_z(\Sigma_{\text{true}}) = E(\mathcal{U} \Sigma_{\text{true}}^{-1} \mathcal{U}^T)$. Hence the result.

References

- Carroll RJ, Maity A, Mammen E, Yu K. Nonparametric additive regression for repeatedly measured data. *Biometrika*. 2009 to appear.
- Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, Wacholder S. Powerful multi-locus tests for genetic association in the presence of gene-gene and gene-environment interactions. *Am. J. Human Genetics* 2006;79:1002–1016. [PubMed: 17186459]
- Fan, J.; Gijbels, I. Local Polynomial Modeling and its Applications. London: Chapman and Hall; 1996.
- Härdle, W.; Liang, H.; Gao, J. Partially Linear Models. Heidelberg: Physica Verlag; 2000.
- Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano R, Bingham S, Schoeller DA, Schatzkin A, Carroll RJ. The structure of dietary measurement error: results of the OPEN biomarker study. *Am. J. Epidemiol* 2003;158:14–21.
- Lin X, Carroll RJ. Semiparametric estimation in general repeated measures problems. *J. Royal Stat. Soc., Series B* 2006;68:68–88.
- Lu Z, Lundervold L, Tjøstheim D, Yao Q. Exploring spatial nonlinearity using additive approximation. *Bernoulli* 2007;13:447–472.

- Maity A, Carroll RJ, Mammen E, Chatterjee N. Powerful multi-locus tests for genetic association with semiparametric gene-environment interactions. *J. Royal Stat. Soc., Series B* 2009;71:75–96.
- Maity A, Ma Y, Carroll RJ. Efficient estimation of population-level summaries in general semiparametric regression models with missing response. *J. Am. Stat. Assoc* 2007;102:123–139.
- van de Geer, S. Empirical processes in M-estimation. Cambridge: Cambridge University Press; 2000.
- Wang N. Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* 2003;90:43–52.
- Wang N, Carroll RJ, Lin X. Efficient semiparametric marginal estimation for longitudinal/clustered data. *J. Am. Stat. Assoc* 2005;100:147–157.
- Willett, WC. *Nutritional Epidemiology*. New York, NY: Oxford University Press; 1990.
- Yu K, Mammen E, Park B. Semiparametric additive regression: gains from the additive structure of the infinite dimensional parameter. Preprint. 2009

Figure 1.
Residual plots from parametric polynomial models and semiparametric models with WI and WD estimators.

Figure 2.

OPEN Study data: nonparametric and polynomial function fits for body mass index, left column, and the protein recall, right column, as predictors of the protein biomarker, "Protein B", along with the 95% pointwise confidence interval for the nonparametric fit. The dashed line is the nonparametric fit, the solid line the polynomial fit and dotted lines represent the confidence intervals.

$MSEs \times 10^4$ of the estimators for β based on 500 samples of size 200. LC stands for local constant estimator and LL for local linear. The seven covariance types are described in the text. Method A accounts for additivity and within-cluster correlation. Method B accounts only for additivity. Method C accounts only for within-cluster correlation. Method D accounts for neither additivity nor within-cluster correlation.

Table 1

covariance type	Local Constant				Local Linear			
	A	B	C	D	A	B	C	D
1	2	12	6	35	2	12	6	36
2	9	13	27	36	9	13	28	37
3	15	16	41	42	16	16	42	43
4	4	14	12	36	4	14	12	37
5	2	14	7	34	3	14	7	35
6	6	62	18	172	6	63	19	177
7	31	67	92	164	31	67	95	167

Table 2

The estimates of β from parametric polynomial models: WD means taking into account the within-cluster correlations, while WI ignores these correlations.

	Estimate	std err
Parametric WI	0.2004	0.02854
Parametric WD	0.1719	0.02686
Semiparametric WI	0.1969	0.02882
Semiparametric WD	0.1529	0.02640