Published in final edited form as:

J Stat Plan Inference. 2012 September 1; 142(9): 2599–2606. doi:10.1016/j.jspi.2012.02.027.

A Note on Generalized Functional Linear Model and Its Application

Qi Long^a

^aDepartment of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, U.S.A.

Abstract

Motivated by a biomarker study for colorectal neoplasia, we consider generalized functional linear models where the functional predictors are measured with errors at discrete design points.

Assuming that the true functional predictor and the slope function are smooth, we investigate a two-step estimating procedure where both the true functional predictor and the slope function are estimated through spline smoothing. The operating characteristics of the proposed method are derived; the usefulness of the proposed method is illustrated by a simulation study as well as data analysis for the motivating colorectal neoplasia study.

Keywords

Generalized functional linear model; Measurement errors; Spline smoothing

1. Introduction

In biomedical studies, predictors are often measured from the same subjects repeatedly over time or a certain spatial structure and are therefore of functional nature. In particular, our work here is motivated by a colorectal neoplasia study, where the goal is to associate a subject's disease status with gene biomarkers whose expression levels in terms of protein contents are measured along the length of colon crypts, a microscopic structure in the human colon mucosa (Daniel et al, 2009). As Figure 1 in Daniel et al. (2009) shows, the distribution of gene expression levels can be measured from the base to the apex of a semi-crypt, which forms a natural one dimensional spatial structure.

Studies like this can be naturally modeled using the generalized functional linear regression (GFLM, for short), where the dependence of a scalar outcome of interest, y , on a functional predictor, $x(\cdot)$ is characterized by a conditional density from the exponential family:

$$f(y|x) = \exp \{ [y\eta(x) - b(\eta(x))] / a(\phi) + c(y, \phi) \}, \quad (1)$$

where

$$\eta(x) = \alpha_0 + \int_{\tau} x(t) \beta_0(t) dt \quad (2)$$

is the natural parameter, ϕ is a nuisance parameter, and $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are specific smooth functions. In parallel to the classical generalized linear models (McCullagh and Nelder, 1989), α_0 and $\beta_0(\cdot)$ are referred to as the intercept and slope function, respectively, in GFLM. The goal is to estimate the intercept and slope function from n iid copies of the pair: $(x_f(\cdot), y_i)$, $i = 1, \dots, n$. Oftentimes, the task is further complicated by the lack of direct observations of $x_f(\cdot)$. Rather, one only has access to noisy observations of $x_f(\cdot)$ at discrete design points:

$$z_{ij} = x_i(t_{ij}) + \epsilon_{ij}, \quad j = 1, \dots, m_i, \quad (3)$$

where the measurement error ϵ_{ij} is independent of the random function $x_f(\cdot)$.

Due to its wide applications, models with functional predictors have drawn much attention in recent years including functional linear models (FLM, for short) (Cardot et al, 2003, 2007; Li and Hsing, 2006; Yao et al., 2005; Crambes et al., 2009; Yuan and Cai, 2010) and GFLM (Cardot and Sarda, 2005; Muller and Stadtmüller, 2005). Most of existing approaches (Cardot et al., 2003; Cardot and Sarda, 2005; Muller and Stadtmüller, 2005; Yuan and Cai, 2010) assume that direct observations of $x(\cdot)$ are available, i.e., $x(\cdot)$ is fully observed without errors. This restriction is lifted only in several recent studies of the standard FLM (Cardot et al., 2007; Li and Hsing, 2006; Yao et al., 2005; Crambes et al, 2009) where $x(\cdot)$ is not fully observed and is measured with errors, and these models can be viewed as a special case of Model (1); most of these current work (Yao et al, 2005; Crambes et al., 2009) exploits the fact that there is a closed-form solution for estimating $\beta_0(t)$ in FLM when $x(\cdot)$ is fully observed without errors. Since such closed-form solution is not available for GFLM when $x(\cdot)$ is fully observed without errors, it is not trivial to extend these results developed for FLM to the more general GFLM and it is also unclear to what extent the existing results for FLM apply to GFLM.

To address our problem of interest, we adopt an approach similar to Li and Hsing (2006); specifically, we investigate a two-step estimating procedure, where both the functional predictor $x_f(\cdot)$ and the slope function $\beta_0(\cdot)$ are estimated through spline smoothing. We provide the details of the estimating procedure in Section 2 and study its operating characteristics in Section 3. In Section 4, we conduct a small simulation study to evaluate the finite sample performance, and illustrate the proposed approach using a colorectal cancer study. Finally, we make some conclusion remarks in Section 5. An outline of the proofs for the main theoretical results is given in Appendix.

2. Estimation

To fix idea, we assume that the true slope function $\beta_0(\cdot)$ belongs to the q th order periodic Sobolev space:

$$W_{per,2}^q = \{f: f, f^{(1)}, \dots, f^{(q-1)} \text{ absolutely continuous and } g^{(k)}(0) = g^{(k)}(1) \text{ for } 0 \leq k \leq q-1, f^{(q)} \in L_2[0, 1]\}.$$

Furthermore, we shall assume that the functional predictor $x(\cdot)$ belongs to the same functional space almost surely with $E(\|x^{[q]}\|^2) < \infty$. Throughout, we denote by $\|\cdot\|$ the usual L_2 norm, and by $\langle \cdot, \cdot \rangle$ the usual L_2 inner product. We note that the q th order Sobolev space

$$W_2^q = \{f: f, f^{(1)}, \dots, f^{(q-1)} \text{ absolutely continuous, } f^{(q)} \in L_2[0, 1]\}$$

is the sum of two spaces, one is $W_{per,2}^q$ and the other is the space spanned by the first $q-1$ polynomial basis functions. Hence, when x and β_0 belong to W_2^q , our results also hold. Hence, this setting is suitable for most applications including the aforementioned colorectal cancer study, and is commonly adopted in the previous studies of functional linear regression (see, e.g., Li and Hsing (2006)).

To motivate our method, we consider first the situation where the functional predictor x_i 's are available. It is evident that in this case, the negative log-likelihood can be expressed as

$$L(\alpha, \beta) = -\frac{1}{n} \sum_{i=1}^n \{y_i \eta(x_i) - b(\eta(x_i))\} \quad (4)$$

up to terms not depending on α and β . The intercept and slope function can then be estimated through penalization:

$$(\hat{\alpha}, \hat{\beta}) = \arg \min \left(L(\alpha, \beta) + \frac{\lambda}{2} J(\beta) \right) \quad (5)$$

where $\lambda \geq 0$ is a tuning parameter, and J is a penalty functional. In particular, we consider the following popular choice of the penalty functional:

$$J(\beta) = \int_0^1 (\beta^{[q]}(t))^2 dt.$$

Proposition 1

Suppose that for $L(\alpha, \beta)$ is continuous and convex with respect to its second argument. Then $\hat{\alpha}$ and $\hat{\beta}$ are uniquely defined if and only if $\text{var}(\int_{\tau} x(t) dt) > 0$.

Proposition 1 can be readily proved and it indicates that this procedure indeed leads to valid estimates when $x_i(\cdot)$ is directly observed. In this case, the estimate of the slope function can be obtained by extending the method proposed in Yuan and Cai (2010), and is not the focus of this article.

We now consider the case when $x_i(\cdot)$ is not observable and only z_{ij} 's as given in (3) are observed. To use the procedure described above, we first construct an estimate of $x_i(\cdot)$, say, $x_i^*(\cdot)$. In particular, we propose to estimate x_i by means of penalized regression splines using the first $2K+1$ Fourier basis functions:

$$x_i^*(\cdot) = \arg \min \left(\frac{1}{m_i} \sum_{j=1}^{m_i} \{z_{ij} - x(t_{ij})\}^2 + \frac{\lambda_i}{2} \int_0^1 (x^{[q]}(t))^2 dt \right) \quad (6)$$

where $\lambda_j = 0$ is a tuning parameter. It can be readily shown that $x_i^* = \sum_{k=1}^{2K+1} \widehat{a}_k \phi_k(\cdot)$. Once $x_i^*, i=1, \dots, n$, we shall use them as if they were the true functional predictor and estimate the intercept and slope function by

$$(\widehat{\alpha}^*, \widehat{\beta}^*) = \arg \min \left(L^*(\alpha, \beta) + \frac{\lambda^*}{2} J(\beta) \right) \quad (7)$$

where

$$L^*(\alpha, \beta) = -\frac{1}{n} \sum_{i=1}^n \{y_i \eta(x_i^*) - b(\eta(x_i^*))\}. \quad (8)$$

Using this two-step estimation approach, the form of the estimate $\widehat{\beta}^*$ is provided as follows:

Proposition 2

Suppose $\{\phi_k: k=1, 2, \dots, \infty\}$ is the set of Fourier basis functions in $\mathcal{H}$, and given

$$x_i^* = \sum_{k=1}^{2K+1} \widehat{a}_k \phi_k(\cdot), \text{ then } \widehat{\beta}^* = \sum_{k=1}^{2K+1} \widehat{b}_k \phi_k(\cdot).$$

There is a well-known connection between penalized splines models and linear mixed models; as a result, in the first step the penalized spline model (6) can be represented using an equivalent linear mixed model and the estimation can then be implemented using routines in existing statistical software such as R, for which the tuning parameter can be chosen by REML. In the second step, applying Proposition 2, the estimates as defined in (7) are equivalent to those of a generalized ridge regression model, e.g., a logistic ridge regression for binary y_i ; for these models, estimation can also be implemented using routines in existing software and the tuning parameter (λ) can be selected using cross-validation techniques (Wahba, 1990), which are aimed to maximize the log-likelihood.

3. Operating Characteristics

We now turn to the operating characteristics of the proposed estimator. In particular, we establish the convergence rate of $\widehat{\beta}^*$ in terms of a semi-norm V , defined as follows

$$V(f) = \int \langle x, f \rangle^2 \nu_0(x) d\omega(x) = E\{\langle x, f \rangle^2 \nu_0(x)\}$$

for any $f \in W_{per,2}^q$. For simplicity, we shall assume in this section that $\alpha = 0$. The discussion and results can be easily extended to handle non-vanishing intercept as well. We shall also assume that $m_1 = \dots = m_n = m$ for brevity. We note that the rate of convergence of the estimate holds so long as there exist constant $C_1, C_2 > 0$ such that $C_1 m \leq m_i \leq C_2 m$ for all $i = 1, \dots, n$.

Throughout, our main results rely on the following main regularity condition:

Condition 1

$$E\|x\|^4 < \infty.$$

We are now ready to state our main result.

Theorem 1

When only z_{ij} 's are observed, $V(\hat{\beta}^* - \beta_0) = O_p(n^{-\frac{2q}{2q+1}} + m^{-\frac{2q}{2q+1}})$ if λ^* is of the order $n^{-\frac{2q}{2q+1}}$ and λ_i s are of the order $m^{-\frac{2q}{2q+1}}$.

We note that $V(\hat{\beta}^* - \beta_0)$ captures the prediction error in the linear predictor of a GLM. The result in Theorem 1 can be established in the following steps. We first obtain a quadratic approximation of the log-likelihood and a resulting intermediate estimator $(\hat{\beta}_1^*)$ and subsequently show that $V^*(\hat{\beta}_1^* - \beta_0)$ and $V^*(\hat{\beta}^* - \beta_1^*)$ achieve the desired convergence rate, where V^* is the seminorm based on estimated $x(\cdot)$, namely, $x^*(\cdot)$, i.e., V^* is defined as V with $x(\cdot)$ replaced with $x^*(\cdot)$. Lastly, we obtain the convergence rate of V^* to V and then Theorem 1 follows. The technical details are provided in Appendix. In the case of known x_i 's, the following corollary can be readily shown using Theorem 1:

Corollary 1

When $x_i(\cdot)$'s are available, $V(\hat{\beta} - \beta_0) = O_p(n^{-\frac{2q}{2q+1}})$ if λ is of the order $n^{-\frac{2q}{2q+1}}$.

4. Numerical Results

4.1. Simulation

We conduct simulation studies to examine the finite sample performance of the proposed estimation approach in a setting where a single functional predictor is measured with errors at discrete design points. The true functional predictor $x_i(i = 1, \dots, n)$ is generated as follows

$$x_i(t) = a_{i,0} + \sum_{k=1}^{\infty} a_{i,k} \sqrt{2} \cos(2k\pi t) / (k+0.5)^2 + \sum_{k=1}^{\infty} a'_{i,k} \sqrt{2} \sin(2k\pi t) / (k+0.5)^2,$$

where $a_{i,k}, a'_{i,k} \sim \text{Normal}(0,1)$. The true slope function, $\beta_0(\cdot)$, is assumed to admit the following expansion

$$\beta_0(t) = b_0 + \sum_{k=1}^{\infty} b_{1,k} \sqrt{2} \cos(2k\pi t) + \sum_{k=1}^{\infty} b_{2,k} \sqrt{2} \sin(2k\pi t),$$

where $b_0 = 2$, $b_{1,1} = 1$, $b_{2,1} = 1$, $b_{1,2} = 0.8$, $b_{2,2} = 0.7$, $b_{1,3} = 0.5$, $b_{2,3} = -0.6$, $b_{1,4} = 0.4$, $b_{2,4} = -0.4$, and $b_{1,k} = b_{2,k} = 1/k^3$ for $k \geq 5$. y_{ij} s then generated from a Bernoulli distribution with the probability of success, π_i , satisfying

$$\text{logit}(\pi_i) = \alpha + \int x_i(t) \beta_0(t) dt = \alpha + a_{i,0} b_0 + \sum_{k=1}^{\infty} (a_{i,k} b_{1,k} + a'_{i,k} b_{2,k}) / (k+0.5)^2$$

where $\alpha = 0.5$. Under this simulation setup, the Bayes risk under 0-1 loss function defined as $\sum_i \min\{\pi_i, 1 - \pi_i\}$ is 0.210. The functional predictor $x_i(t)$ is not directly observed, and we instead observe

$$z_i(t_{ij}) = x_i(t_{ij}) + \epsilon_i(t_{ij}) \quad j=1, \dots, m$$

at m equally spaced design points in $(0,1)$, where $\epsilon_i(t_{ij})$ are i.i.d. Normal $(0, \sigma_z^2)$ with $\sigma_z^2=1$. For each simulation run, a training data set is used to obtain parameter estimates and a testing data set is used to evaluate the convergence criteria and prediction error. For each Monte Carlo training data set and test data set, n observations were generated; we considered different sample sizes, that is, $n = 100, 400$, and $m = 30, 100$.

The proposed approach discussed in Section 2 is used for the analysis of each Monte Carlo training data set. In the first step of the data analysis, $m/2$ basis functions are used to model x_i . In the second step, a functional logistic regression model is fit to obtain the estimate of β . The tuning parameter λ^* is chosen using a k -fold ($k = 50$) cross-validation approach, which maximizes the log-likelihood function for a logistic regression. After obtaining $\hat{\beta}^*(\cdot)$, the performance of the proposed estimator is evaluated based on three measures, namely, $V^*(\hat{\beta}^* - \beta_0)$, $V(\hat{\beta}^* - \beta_0)$, and the prediction error based on the prediction rule, $\hat{y} = I(\hat{\pi} \geq 0.5)$, where $I(\cdot)$ is the indicator function. These measures are computed using an independent Monte Carlo test data set. Of note, to compute $V^*(\hat{\beta} - \beta_0)$, x^* also needs to be estimated in the test data set.

The results are summarized over 1000 independent replications and provided in Table .1. Our simulation results show that the performance of the proposed method in finite sample is satisfactory, in particular, in terms of the prediction error. In general, as n or m increases, $V^*(\hat{\beta}^* - \beta_0)$, $V(\hat{\beta}^* - \beta_0)$, and prediction error decrease, and the difference between $V^*(\hat{\beta}^* - \beta_0)$ and $V(\hat{\beta}^* - \beta_0)$ decreases. When $n = 400$ and $m = 100$, the prediction error is very close to the true Bayes risk. The impact of increasing m and n is different on $V^*(\hat{\beta}^* - \beta_0)$ and $V(\hat{\beta}^* - \beta_0)$ compared to its impact on the prediction error. When $n = 100$ and $m = 100$, the prediction error is similar but $V^*(\hat{\beta}^* - \beta_0)$ and $V(\hat{\beta}^* - \beta_0)$ are considerably larger, compared to when $n = 400$ and $m = 30$. Our results seems to suggest that the number of observations n has a larger impact on the performance evaluated by $V^*(\hat{\beta}^* - \beta_0)$ and $V(\hat{\beta}^* - \beta_0)$.

4.2. Application to the Colorectal Neoplasia Study

We illustrate the proposed method using data from a colorectal neoplasia study, the Markers of Adenomatous Polyps II (MAP II) study. The MAP II study is a case-control study to evaluate biomarkers of risk for colorectal cancer (Daniel et al., 2009). If the participants are found with adenoma through colonoscopy, they are considered “cases” (at higher risk for colorectal cancer) and are denoted by $y = 1$; if the participants are found with no adenoma, they are considered “controls” (at lower risk for colorectal cancer) and are denoted by $y = 0$. Tissue samples from patients are first immunohistochemically processed to stain protein biomarkers of interest; then the biomarker levels are measured in terms of optical density along the length of colon crypts, which are microscopic structures in the human colon mucosa. It is well known that measurement errors are present in this process. The entire crypt length is standardized into 50 segments with 1 denoting the first segment at the base and 50 denoting the apical segment, and the optical density is averaged within each segment (see Figure 1 in Daniel et al. (2009)). In the MAP II study, multiple crypts are sampled and measured for each subject, and due to missing data only segments 2-49 are used in our data analysis. For each subject, we use the mean optical density over multiple crypts from the same subject as the observed $z_A(t_j)$ at $t_j = j/50$ ($j = 2, \dots, 49$). Following Prentice and Pyke (1979), it can be readily shown that case-control studies such as the MAP II study can be

analyzed using the propose method as if they were prospective studies. While the interpretation for α is no longer valid, the interpretation of $\beta(\cdot)$ still holds.

For illustration purpose, we only use one functional predictor in Model (2), namely, the expression levels of APC, a known tumor suppressor gene. A total of 42 cases and 45 controls are used in our data analysis. The proposed method is used to fit GFLM. A logistic model is posited for the outcome of interest, risk for colorectal cancer, that is, $\text{logit}[P(y_i = 1/x_i(\cdot))] = \alpha + \int \beta(t)x_i(t)dt$ where $x_i(t)$ is the underlying true distribution of APC expression levels along the length of crypts for subject i and is not directly observed. We compare the results using different number of Fourier basis functions and the results are very similar, which indicates that this data analysis is not sensitive to the number of basis functions. The results are reported for the data analysis using 47 Fourier basis functions.

Figure .1 provides a sample of the observed biomarker (APC) measurements ($z_i(t_j)$) and the estimated curves ($\hat{x}_i^*$) in the first step. It shows that the penalized spline estimates fit the data well, and it also shows that the controls (the second row in Figure .1) tend to have higher APC values than the cases (the first row in Figure .1) along the length of colon crypts. Figure .2 displays the estimated slope function $\hat{\beta}^*(\cdot)$ from the second step. In general, the negative part of the slope function indicates that lower APC expression is associated with higher risk for colorectal cancer ($y = 1$) in that region, and the positive part of the slope function indicates that higher APC expression is associated with higher risk for colorectal cancer. A larger absolute value of the slope function indicates that the strength of the association is stronger between APC expression in that part of a crypt and risk for colorectal cancer. Figure .2 shows that the estimated slope function is mostly negative except for the regions near the base and apex of crypts. Figure .2 also shows that the APC expression around segment location 24 had the largest absolute slope function value and therefore their association with risk for colorectal cancer is likely the strongest, suggesting that the region near this location is important in terms of distinguishing cases from controls and specifically lower values of APC in this region is associated with higher risk for colorectal cancer.

5. Conclusions

In this paper, we investigate a spline smoothing approach to estimate the slope function $\beta_0(\cdot)$ in generalized functional linear models, and we derive the operating characteristics of the proposed estimator of $\beta_0(\cdot)$ in cases where the functional predictor is measured at discrete design points with errors. Our numerical studies showcase the good finite sample performance of the proposed estimator as well as its usefulness in practice. In this research, we use the set of Fourier basis functions and it is straightforward to extend our results using other basis functions.

Acknowledgments

This work was supported in part by PHS Grant UL1 RR025008 from the Clinical and Translational Science Award program, National Institutes of Health, National Center for Research Resources.

Appendix

Proof of Proposition 2

Given $\{\phi_k: k = 1, 2, \dots, \infty\}$ is the set of Fourier basis functions for $\cdot$, then we have

$$\beta = \sum_{k=1}^{2K+1} b_k \phi_k(\cdot) + \sum_{k=2K+2}^{\infty} b_k \phi_k(\cdot). \text{ Given } x^* = \sum_{k=1}^{2K+1} \hat{a}_k \phi_k(\cdot), \text{ then}$$

$$\eta = \alpha \int x^*(t) \beta(t) dt = \alpha + \sum_{k=1}^{2K+1} \hat{a}_k b_k \text{ and therefore } L_n^*(\beta) \text{ in the objective function (6) does not}$$

depend on $\{b_k, k = 2K + 2, \dots, \infty\}$. In addition, one can show

$J(\beta) = \sum_{k=1}^{2K+1} k^{2q} b_k^2 + \sum_{k=2K+2}^{\infty} k^{2q} b_k^2$. It is then obvious that $b_k = 0$ for $k = 2K + 2, \dots, \infty$ for estimates 7. Hence, $\widehat{\beta}^* = \sum_{k=1}^{2K+1} \widehat{b}_k \phi_k$.

Proof of Theorem 1

Let Q denote the upper bound of the convergence rates of $\|x_i^* - x_i\|^2$. Using an argument similar to that for Theorem 2 in Li and Hsing (2006), $Q = O_p\left(m^{-\frac{2q}{2q+1}}\right)$. We first state three lemmas and defer their proofs to the end of Appendix, where β_1^* denotes an intermediate estimator based on a quadratic approximation to the log-likelihood (8) defined in the proofs of Lemmas 1-3.

Lemma 1

As $n(\lambda^*)^{1/q} \rightarrow \infty$ and $\lambda^* \rightarrow 0$, $V^*(\beta_1^* - \beta_0) = O_p\left(n^{-1}(\lambda^*)^{-1/(2q)} + \lambda^* + Q\right)$.

Lemma 2

As $n(\lambda^*)^{1/q} \rightarrow \infty$ and $\lambda \rightarrow 0$, $V^*(\widehat{\beta}^* - \beta_1^*) = O_p\left(n^{-1}(\lambda^*)^{1/(2q)} + \lambda^* + Q\right)$.

We define $V^*(g) = \frac{1}{n} \sum_{i=1}^n \langle g, x_i^* \rangle^2 \nu_0(x_i^*)$ and let F and F^* be the variance-covariance operator that corresponds to the inner product $V(\cdot, \cdot)$ and $V^*(\cdot, \cdot)$, respectively. In other words, $V(f, g) = \langle f, Fg \rangle$ and $V^*(f, g) = \langle f, F^*g \rangle$. F depends on $x(\cdot)$'s and F^* depends on $x_i^*(\cdot)$'s.

Lemma 3

$E\left(\|F^* - F\|_{\mathcal{H}}^2\right) = O\left(n^{-1} + Q\right)$, where $\|\cdot\|$ is the Hilbert-Schmidt norm of the operator.

Using an argument similar to that for Theorem 5 in Li and Hsing (2006), it follows immediately from Lemma 1, 2 and 3 that $V(\widehat{\beta}^* - \beta_0) = O_p\left(n^{-1}(\lambda^*)^{-1/(2q)} + \lambda^* + Q\right)$. Theorem 1 follows immediately.

Proofs of Lemmas 1-3

Let $\Phi_K = (\phi_1, \dots, \phi_{2K+1})^T$ be the set of first $2K + 1$ ortho-normal Fourier basis functions in $W_{2,per}^q$. Then we have $x_i^*(t) = \Phi_K^T(t) w_i$, where $w_i = (w_{i1}, \dots, w_{iK})^T$ is the vector of corresponding coefficients. w_i are estimated using the observed data $z_i = (z_i(t_1), \dots, z_i(t_{im}))^T$ and the proposed penalized regression spline method. For a function g in $W_{2,per}^q$, we denote by $\underline{g}$ as the vector of coefficients of the projection of g onto the space spanned by Φ_K , that is, the projection of g is $\Phi_K^T \underline{g}$.

Using a technique similar to that in Gu and Qiu (1993), we consider a quadratic approximation of L^* as follows

$$L_1^*(\beta) = -\frac{1}{n} \sum_{i=1}^n \left\{ y_i \eta \left[x_i^*(\cdot); \beta \right] - \mu_0 \left[x_i^*(\cdot) \right] \eta \left[x_i^*(\cdot); \beta \right] \right\} + \frac{1}{2} V^*(\beta - \beta_0)$$

where $V^*(g) = \sum_{i=1}^n \langle g, x_i^* \rangle^2 v_0(x_i^*)$, and the corresponding objective function as follows

$$L_1^* + \frac{\lambda^*}{2} J(\beta) = -\frac{1}{n} \sum_{i=1}^n \{y_i \eta[x_i^*(\cdot); \beta] - \mu_0[x_i^*(\cdot)] \eta[x_i^*(\cdot); \beta]\} + \frac{1}{2} V^*(\beta - \beta_0) + \frac{\lambda^*}{2} J(\beta) \quad (.1)$$

It is straightforward to show that $V^*(g) = g^T \Omega g$ where $\Omega = \frac{1}{n} \sum_{i=1}^n w_i w_i^T v_0(x_i^*)$. We note that Ω depends on x_i^* . Since $\mathcal{K}(g) = \int (D^q g)^2$, it can be shown that the minimizer of the quadratic functional (.1) takes the form of $\beta_1^* = \Phi_k^T \underline{\beta}_1^*$, and hence $J(\beta_1^*) = \underline{\beta}_1^{*T} D \underline{\beta}_1^*$, where

$$D = \text{diag} \{0, (2\pi)^{2q}, (2\pi)^{2q}, (4\pi)^{2q}, (4\pi)^{2q}, \dots, (K\pi)^{2q}, (K\pi)^{2q}\}.$$

The estimator x^* proposed in Li and Hsing (2006) is a special case, that is, a penalized

spline estimate with $w_i = \frac{1}{m_i} (\frac{1}{m_i} \Phi_i^T \Phi_i + \lambda^* D)^{-1} \Phi_i^T z_i$, where $\Phi_i = \{\varphi_k(t_{ij})\}$ ($j = 1, \dots, m_i$ and $k = 1, \dots, 2K+1$).

Then we can rewrite the objective function (.1) as

$$-\frac{1}{n} r^T W \underline{\beta}_1^* + \frac{1}{2} (\underline{\beta}_1^* - \underline{\beta}_0)^T \Omega (\underline{\beta}_1^* - \underline{\beta}_0) + \frac{1}{2} \lambda^* \underline{\beta}_1^{*T} D \underline{\beta}_1^*.$$

where $W = (w_1, w_2, \dots, w_n)^T$ and $r = [y_1 - \mu_0(x_1^*), y_2 - \mu_0(x_2^*), \dots, y_n - \mu_0(x_n^*)]^T$. Since $\underline{\beta}_1^*$ minimize (.1), some algebra can show

$$\underline{\beta}_1^* = (\Omega + \lambda^* D)^{-1} \left(-\frac{1}{n} W^T (r_1 + r_2) + \Omega \underline{\beta}_0 \right)$$

where $r_1 = [y_1 - \mu_0(x_1), y_2 - \mu_0(x_2), \dots, y_n - \mu_0(x_n)]^T$ and

$r_2 = [\mu_0(x_1) - \mu_0(x_1^*), \mu_0(x_2) - \mu_0(x_2^*), \dots, \mu_0(x_n) - \mu_0(x_n^*)]^T$. We now proceed to prove Lemmas 1-3.

Proof of Lemma 1

We first establish the convergence rates of $E(V^*(\beta_1^* - \beta_0))$ and $E(J(\beta_1^* - \beta_0))$, where $E(\cdot)$ is taken conditional on the observed curves $z_k(t_{ij})$. Some algebra can show

$$V^*(\beta_1^* - \beta_0) = V^* \left(\Phi_k^T \left[(\Omega + \lambda^* D)^{-1} \left\{ -\frac{1}{n} W^T (r_1 + r_2) - \lambda^* D \underline{\beta}_0 \right\} \right] \right) \leq T_1 + T_2 + T_3$$

where $T_1 = V^* (\Phi_K^T (\Omega + \lambda^* D)^{-1} \{-\frac{1}{n} W^T r_1\})$, $T_2 = V^* (\Phi_K^T (\Omega + \lambda^* D)^{-1} \lambda^* D \beta_0)$, and $T_3 = V^* (\Phi_K^T (\Omega + \lambda^* D)^{-1} \{-\frac{1}{n} W^T r_2\})$.

We consider $E(T_1)$ first,

$$\begin{aligned} E(T_1) &= \frac{1}{n^2} E[r_1^T W (\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} W^T r_1] \\ &= \frac{1}{n^2} Tr\{(\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} W^T E[r_1 r_1^T] W\} \\ &= \frac{1}{n^2} Tr\left\{(\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} \sum_{i=1}^n w_i w_i^T v_0(x_i)\right\} \\ &= \frac{1}{n^2} Tr\left\{(\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} \left[\sum_{i=1}^n w_i w_i^T v_0(x_i^*) + \sum_{i=1}^n w_i w_i^T v_0(x_i^*) \left(\frac{v_0(x_i)}{v_0(x_i^*)} - 1\right)\right]\right\} \\ &= \frac{1}{n} Tr\left\{(\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} [\Omega + \Omega o_p(1)]\right\} \\ &= T_{11} + T_{11} o_p(1) \end{aligned}$$

where $T_{11} = \frac{1}{n} Tr\{(\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} \Omega\}$. Let $\nu_1 \dots \nu_{2K+1}$ be the $2K+1$ eigenvalues of Ω , $d_1 \dots d_{2K+1}$ be the eigenvalues of D , $\xi_1 \dots \xi_{2K+1}$ be the eigenvalues of $\Omega + \lambda^* D$, and $\psi_1 \dots \psi_{2K+1}$ be the eigenvalues of $(\Omega + \lambda^* D)^{-1} \Omega$. We know that $d_1 = 0$, and $d_k = (Lk/2J2\pi)^{2q}$ for $k > 1$. It is straightforward to show that Ω is positive semi-definite, and hence $\nu_k \geq 0$. Applying Theorems 1 and 7 of Merikoski and Kumar (2004), we then have $\xi_k \leq \lambda^* d_k + \nu_1$, $1/\xi_k \leq 1/(\lambda^* d_k + \nu_1) \leq 1/(\lambda^* d_k)$, and $\psi_k \leq \nu_k \xi_k \leq \nu_{2K+1}/(\lambda^* d_k)$. It is also obvious that $\psi_k \leq 1$. Consequently,

$$E(T_{11}) = \frac{1}{n} \sum_{k=1}^{2K+1} \psi_k^2 \leq O_p\left(n^{-1} (\lambda^*)^{-1/(2q)} \nu_{2K+1}^2\right),$$

which holds even as $K \rightarrow \infty$. Under Condition 1, it is straightforward to show that ν_{2K+1} is bounded from above for all K and it follows that

$$E(T_1) = O_p\left(n^{-1} (\lambda^*)^{-1/(2q)}\right). \quad (2)$$

We now consider $E(T_2)$,

$$\begin{aligned} E(T_2) &= (\lambda^*)^2 E\left[\beta_0^T D' (\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} D \beta_0\right] \\ &= \lambda^* \times tr\left\{(\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} (\lambda^* D) (\beta_0 \beta_0^T D')\right\} \\ &\leq C \lambda^* \times tr\left\{(\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} (\lambda^* D)\right\} \\ &= C \lambda^* \times \left[tr\left\{(\Omega + \lambda^* D)^{-1} \Omega\right\} - tr\left\{(\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} \Omega\right\}\right] \\ &= O_p(\lambda^*) \end{aligned}$$

where C is a positive constant, the first inequality is due to the fact that $\mathcal{K}(\beta_0)$ is bounded, and the last equality can be proved using similar arguments for bounding $E(T_1)$. Hence for all K

$$E(T_2) = O_p(\lambda^*). \quad (.3)$$

Finally, we consider $E(T_3)$. Note that the i th element of r_2 is

$\mu_0(x_i) - \mu_0(x_i^*) \approx v_0(x_i^*) \langle x_i - x_i^*, \beta_0 \rangle$. We further define

$$r_3 = \left(v_0(x_1^*)^{1/2} \langle x_1 - x_1^*, \beta_0 \rangle, \dots, v_0(x_n^*)^{1/2} \langle x_n - x_n^*, \beta_0 \rangle \right)^T,$$

$$r_4 = \left(o_p(1) v_0(x_1^*)^{1/2} \langle x_1 - x_1^*, \beta_0 \rangle, \dots, o_p(1) v_0(x_n^*)^{1/2} \langle x_n - x_n^*, \beta_0 \rangle \right)^T \text{ and}$$

$$W_1 = \left(w_1 v_0(x_1^*)^{1/2}, \dots, w_n v_0(x_n^*)^{1/2} \right)^T. \text{ Then}$$

$$E(T_3) = \frac{1}{n^2} E \left[r_3^T W_1 (\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} W_1^T r_3 \right] + \frac{1}{n^2} E \left[r_4^T W_1 (\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} W_1^T r_4 \right] = T_{31} + T_{32}.$$

$\frac{1}{n} W_1 (\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} W_1^T$ and $(\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} \frac{1}{n} W_1^T W_1$ have the same set of nonzero eigenvalues, and $\frac{1}{n} W_1^T W_1 = \Omega$ and the eigenvalues of $[(\Omega + \lambda^* D)^{-1} \Omega]^2$ are bounded by 1. Thus, the eigenvalues of $W_1 (\Omega + \lambda^* D)^{-1} \Omega (\Omega + \lambda^* D)^{-1} W_1^T$ are bounded by 1. Then by the definition of the largest eigenvalue, we have

$$E(T_{31}) \leq \frac{1}{n} E(r_3^T r_3) \leq \frac{\|\beta_0\|^2}{n} \sum_{i=1}^n E \left[\|x_i^* - x_i\|^2 v_0(x_i^*) \right] = O_p(Q)$$

Similarly, it can be shown that $E(T_{32}) = o_p(1) O_p(Q)$. Hence,

$$E(T_3) = O_p(Q). \quad (.4)$$

Combining (.2), (.3) and (.4), Lemma 1 follows.

Proof of Lemma 2

The key idea is to show that $V^*(\hat{\beta} - \beta_1)$ is bounded by $CV^*(\beta_1 - \beta_0)$, where C is a positive constant.

We define $B_{f,g}(\alpha) = L^*(f + \alpha g) + (\lambda^*/2) J(f + \alpha h)$ and

$C_{f,g}(\alpha) = L_1^*(f + \alpha g) + (\lambda^*/2) J(f + \alpha h)$. Taking the first derivative w.r.t. α we have

$$0 = \dot{B}_{\hat{\beta}^*, \hat{\beta}^* - \beta_1^*}(0) = -\frac{1}{n} \sum_i \left[y_i \langle \hat{\beta}^* - \beta_1^*, x_i^* \rangle - \hat{\mu}(x_i^*) \langle \hat{\beta}^* - \beta_1^*, x_i^* \rangle \right] + \lambda^* J(\hat{\beta}^*, \hat{\beta}^* - \beta_1^*), \quad (.5)$$

$$0 = \dot{C}_{\hat{\beta}^*, \hat{\beta}^* - \beta_1^*}(0) = -\frac{1}{n} \sum_i \left[y_i \langle \hat{\beta}^* - \beta_1^*, x_i^* \rangle - \mu_0(x_i^*) \langle \hat{\beta}^* - \beta_1^*, x_i^* \rangle \right] + V^*(\beta_1^* - \beta_0, \hat{\beta}^* - \beta_1^*) + \lambda^* J(\beta_1^*, \hat{\beta}^* - \beta_1^*).$$

Equating (.5) and (.6) and rearranging terms, we have

$$\frac{1}{n} \sum_{i=1}^n [\widehat{\mu}(x_i^*) - \mu_1(x_i^*)] \langle \widehat{\beta}^* - \beta_1^*, x_i^* \rangle + \lambda^* J(\widehat{\beta}^* - \beta_1^*) = V^*(\beta_1^* - \beta_0, \widehat{\beta}^* - \beta_1^*) - \frac{1}{n} \sum_{i=1}^n [\mu_1(x_i^*) - \mu_0(x_i^*)] \langle \widehat{\beta}^* - \beta_1^*, x_i^* \rangle,$$

where $\widehat{\mu}(x_i^*) = E(y|x_i^*; \widehat{\beta}^*)$ and $\mu_1(x_i^*) = E(y|x_i^*; \beta_1^*)$. Assuming the following condition holds,

Condition 2

For β in a convex set B_0 around β_0 containing β_1^* and $\widehat{\beta}^*$, there exist c_1 and $c_2 \in (0, \infty)$ such that $c_1 v_0(x^*) \leq v(x^*) \leq c_2 v_0(x^*)$ uniformly for x^* in $\cdot$.

Then, we have

$$c_1 \frac{1}{n} \sum_{i=1}^n \langle \widehat{\beta}^* - \beta_1^*, x_i^* \rangle^2 v_0(x_i^*) \leq \frac{1}{n} \sum_{i=1}^n [\widehat{\mu}(x_i^*) - \mu_1(x_i^*)] \langle \widehat{\beta}^* - \beta_1^*, x_i^* \rangle,$$

$$c_2 \frac{1}{n} \sum_{i=1}^n \langle \beta_1^* - \beta_0, x_i^* \rangle^2 v_0(x_i^*) \geq \frac{1}{n} \sum_{i=1}^n [\mu_1(x_i^*) - \mu_0(x_i^*)] \langle \widehat{\beta}^* - \beta_1^*, x_i^* \rangle.$$

It follows from (.7) that

$$(c_1 V^* + \lambda^* J)(\widehat{\beta}^* - \beta_1^*) \leq |1 - c_2| V^*(\beta_1^* - \beta_0, \widehat{\beta}^* - \beta_1^*) \leq |1 - c_2| V^*(\beta_1^* - \beta_0)^{1/2} V^*(\widehat{\beta}^* - \beta_1^*)^{1/2},$$

which implies that $V^*(\widehat{\beta}^* - \beta_1^*)$ is bounded by $CV^*(\beta_1^* - \beta_0)$. Together with Lemma 1, Lemma 2 follows immediately.

Proof of Lemma 3

By an argument similar to that for Theorem 4 in Li and Hsing (2006), it can be verified that Lemma 3 holds for $x_i^*(\cdot)$.

References

- Cardot H, Crambes C, Kneip A, Sarda P. Smoothing spline estimators in functional linear regression with errors-in-variables. Computational Statistics and Data Analysis. 2007; 51:4832–48.
- Cardot H, Ferraty F, Sarda P. Spline estimators for the functional linear model. Statistica Sinica. 2003; 13:571–91.
- Cardot H, Sarda P. Estimation in generalized linear models for functional data via penalized likelihood. Journal of Multivariate Analysis. 2005; 92:24–41.
- Crambes C, Kneip A, Sarda P. Smoothing splines estimators for functional linear regression. The Annals of Statistics. 2009; 37:35–72.
- Daniel C, Bostick R, Flanders W, Long Q, Fedirko V, Sidelnikov E, Seabrook M. Tgf- α expression as a potential biomarker of risk within the normal-appearing colorectal mucosa of patients with and without incident sporadic adenoma. Cancer Epidemiology Biomarkers & Prevention. 2009; 18:65–73.
- Gu C, Qiu C. Smoothing spline density estimation: theory. The Annals of Statistics. 1993; 21:217–34.

- Li Y, Hsing T. On rates of convergence in functional linear regression. *Journal of Multivariate Analysis*. 2006; 98:1782–1804.
- McCullagh, P.; Nelder, JA. *Generalized Linear Models*. 2nd. New York: Chapman & Hall/CRC; 1989.
- Merikoski JK, Kumar R. Inequalities for spreads of matrix sums and products. *Applied Mathematics E-Notes*. 2004; 4:150–159.
- Muller HG, Stadtmüller U. Generalized functional linear models. *The Annals of Statistics*. 2005; 33:774–805.
- Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika*. 1979; 66:403–11.
- Wahba, G. *Spline Models for Observational Data*. Philadelphia: SIAM; 1990.
- Yao F, Muller H, Wang J. Functional linear regression analysis for longitudinal data. *The Annals of Statistics*. 2005; 33:2873–2903.
- Yuan M, Cai T. A reproducing kernel hilbert space approach to functional linear regression. *Annals of Statistics*. 2010; 38(6):3412–3444.

Figure .1. Observed biomarker (APC) values (circle) along the length of crypts for the MAP II study and their estimates (solid line) using the first step of the proposed method. The first row displays two samples from cases and the second row displays two samples from controls.

Figure .2. Estimated slope function using the proposed method

Table .1

Simulation study. The performance of the proposed estimator ($\hat{\beta}^*$) is evaluated by $V^*(\hat{\beta}^* - \beta_0)$, $V(\hat{\beta}^* - \beta_0)$, and the prediction error for different n (the number of observations) and m (the number of repeated measurements of the functional predictor for each observation). The Bayes risk is 0.210.

n	m	$V^*(\hat{\beta}^* - \beta_0)$	$V(\hat{\beta}^* - \beta_0)$	Prediction Error
100	30	0.406	0.378	0.232
	100	0.360	0.344	0.224
400	30	0.144	0.122	0.223
	100	0.106	0.097	0.215