Published in final edited form as:

Comput Stat Data Anal. 2016 February 1; 94: 317–329. doi:10.1016/j.csda.2015.08.020.

Interaction Models for Functional Regression

JOSEPH USSET[a], **ANA-MARIA STAICU**[b], and **ARNAB MAITY**[b,1]

[a]Kansas University Department of Biostatistics, Kansas City, KS, USA

[b]North Carolina State Department of Statistics, Raleigh, NC, USA

Abstract

A functional regression model with a scalar response and multiple functional predictors is proposed that accommodates two-way interactions in addition to their main effects. The proposed estimation procedure models the main effects using penalized regression splines, and the interaction effect by a tensor product basis. Extensions to generalized linear models and data observed on sparse grids or with measurement error are presented. A hypothesis testing procedure for the functional interaction effect is described. The proposed method can be easily implemented through existing software. Numerical studies show that fitting an additive model in the presence of interaction leads to both poor estimation performance and lost prediction power, while fitting an interaction model where there is in fact no interaction leads to negligible losses. The methodology is illustrated on the AneuRisk65 study data.

Keywords

Functional regression; Hypothesis Testing; Interaction; Spline smoothing

1. Introduction

Functional regression models with scalar response and functional covariate have received a considerable amount of attention in the functional data analysis literature. Perhaps one of the most popular functional regression models is the so called functional linear model (FLM), first introduced by [41]. A typical FLM with a single functional predictor quantifies the effect of the predictor as an inner product between the functional predictor and an unknown coefficient function; see e.g., [27, 40, 18, 6] for general discussions on this type of model. Recently, there has been a lot of interest in functional regression models that relax the linearity assumption used in FLM. For the case of a single functional predictor, current advances in this direction include: purely nonparametric functional regression models (see [13, 18]) and functional partially linear models, where the functional covariate is modeled nonparametrically and other scalar or vector valued covariates are modeled parametrically

[1]Correspondence to: 2311 Stinson Drive, Campus Box 8203, Raleigh, NC, 27695, USA. Tel: 919-597-1937. amaity@ncsu.edu.

The authors report no conflict of interests.

(see e.g., [3, 4, 32, 33], among many others). These models are commonly developed using nonparametric kernel smoothing based-techniques. In the spline smoothing framework, [58] developed spline estimation for a semi-functional linear model, while [36] and [35] developed estimation and testing procedures for functional generalized additive models. Non-linear extensions to the usual FLM include single-index models, where instead of modeling the entire functional covariate nonparametrically, one models a linear index (defined by the inner product of the function with an unknown coefficient function) via an unknown smooth function; see e.g., [30, 1, 8, 17] and references therein. A kernel machine regression based approach to fit a linear functional regression model was proposed by [57]. Recently, [31] developed kNN based estimation procedure for nonparametric functional regression models and provided uniform consistency results.

Applications involving two or more functional covariates are becoming increasingly popular. There are several extensions of the simple FLM that incorporate multiple functional predictors: 1) such as generalized functional linear models [29] for exponential family response variables, 2) penalized functional regression [24], 3) group lasso based variable selection for functional linear models [20], 4) linear functional additive models for time series prediction [22], among many others. Extensions beyond the linear relationship include: functional partially linear models, where some of the functional covariates are modeled are modeled nonparametrically while the rest of the covariates are modeled linearly, see for example [5] and [32], among others. Fully nonparametric functional regression models were recently developed for both continuous and general response variables in [37] and [19], respectively, where each of the functional predictors are modeled using smooth nonparametric functionals. These articles also include development of functional index models with multiple functional predictors. Recently, [23] proposed a partitioned functional single-index model where the domain of functional covariate in partitioned into several smaller interval and separate indices are formed for each interval, and the indices are modeled nonparametrically in an additive fashion. Multivariate functional non-parametric models and additive functional non-parametric models are developed by [3]. There are several resources (such as [6, 18, 27, 40]) that provide extensive discussion on various types of functional regression models; we refer the readers to these resources for further background.

While there is a significant amount of literature available on functional regression with multiple predictors, a common assumption made by all the above mentioned models is that the effects of the functional predictors are additive, that is only the main effects of the individual functional covariates enter the regression model. Thus any interaction between the functional covariates are not taken into account. In general, ignoring such interaction terms may lead to inaccurate and biased estimation of the model parameters which in turn lead to incorrect conclusions. Therefore, development of a functional regression model is needed where one can accommodate both multiple functional predictors as well as interactions among them. In this article, we develop a functional linear interaction model, as well as a penalized spline based estimation procedure for the interaction effect and individual main effects of the functional covariates.

The model we consider is described as follows. Suppose for $i = 1, \ldots, n$, we observe a scalar response Y_i, and independent real-valued, zero-mean, and square integrable random functions $X_{1i}(\cdot)$ and $X_{2i}(\cdot)$ observed without noise, on dense grids. We consider the model

$$E[Y_i|X_{1i}, X_{2i}]=\alpha+\int X_{1i}(s)\beta_1(s)ds+\int X_{2i}(t)\beta_2(t)dt+\int\int X_{1i}(s)X_{2i}(t)\gamma(s,t)dsdt, \quad (1)$$

where α is the overall mean, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are real-valued functions defined on τ_1 and τ_2 respectively, and $\gamma(\cdot, \cdot)$ is a real valued bi-variate function defined on $\tau_1 \times \tau_2$. The unknown functions β_1 and β_2 capture the main effects of the functional covariates, while γ captures the interaction effect. To gain some insight, consider the particular case $\beta_1(\cdot) \equiv \beta_{01}$, $\beta_2(\cdot) \equiv \beta_{02}$, $\gamma(\cdot, \cdot) \equiv \gamma_0$, for scalars β_{01}, β_{02}, and γ_0. This case reduces to the common two-way interaction model, with covariates $\bar{X}_{ji} = X_{ji}(s)\, ds$, which act as a sufficient summaries, X_{ji}, $j = 1, 2$. Thus the proposed model is an extension of the common two-way interaction model from scalar covariates to functional covariates. The denseness of the sampling design and the noise free assumption are made for simplicity and will be relaxed in later sections.

Recently, [55] introduced a class of functional polynomial regression models of which model (1) is a special case; they showed that accounting for a functional interaction effect between depth spectrograms and temperature time series improved prediction of sturgeon spawning rates in the Lower Missouri river. The proposed methodology relies on an orthonormal basis decomposition of the functional covariates and parameter functions, combined with stochastic search variable selection in a fully Bayesian framework. Their approach requires full prior specification of several parameters, along with implementation of an MCMC algorithm for model fitting.

The main contribution of this article is a novel approach for estimation, inference and prediction in a parametric functional linear model that incorporates a two-way interaction. We consider a frequentist view and model the unknown functions using pre-determined spline bases and control their smoothness with quadratic penalization. The inclusion of an interaction term between the functional predictors involves additional computational and modeling challenges. A tensor product basis is used to model the interaction surface; such a choice is particularly attractive as it can automatically handle predictors that are on different scales, allows for flexible smoothing in separate directions of the interaction contour, and easily extends to higher dimensions; see [12] for important early work, see also [16]. The main advantage of our approach is that it can be implemented with readily available software, that accomodates 1) responses from any exponential family, 2) functional covariates observed with error, or on a sparse or dense grid, and 3) produces p-values for individual model components, which include the interaction term. The paper also includes a numerical comparison between the additive and interaction functional models involving scalar response. Our findings can be summarized as follows. When the true model contains an interaction between the functional covariates, as specified in (1), then fitting a simpler additive model [24] leads to biased estimates and low prediction performance compared to fitting a functional interaction model. When the true model contains no interaction effect, then with sufficient sample size, fitting the more complex functional interaction model does not harm the estimation, inference or prediction performance.

The remainder of this paper is as follows. In Section 2, we develop the estimation framework of the model in (1). Section 3 extends the methodology to handle general outcomes or where predictors are measured sparsely or with error; and describes hypothesis testing for interaction. In Section 4, we evaluate our method via a simulation study. In Section 5, we apply the interaction model to the AneuRisk65 data. Sections 6 and 7 discuss implementation and present future directions for research, respectively.

2. Modeling Methodology

2.1. Estimation

We first discuss the case when the response variable is continuous and the covariates are observed on a dense design and without noise. In later sections, we generalize our procedure to accommodate noisy and/or sparely observed predictors as well as generalized response variables. The central idea behind our approach is to model the parameter functions using pre-specified bases and then use a penalized estimation procedure to control smoothness of the estimates.

In this article, we consider basis function decompositions of the parameter functions using known spline bases. Specifically, let $\{\psi_{1k}(s)\}_{k=1}^{K}$ and $\{\psi_{2l}(t)\}_{l=1}^{L}$ be two bases in $\mathscr{L}^2(\tau_1)$ and $\mathscr{L}^2(\tau_2)$ respectively, and furthermore let $\{\varphi_{kl}(s, t) = \psi_{1k}(s)\psi_{2l}(t)\}_{1\le k\le K, 1\le l\le L}$ be the corresponding tensor product basis in $\mathscr{L}^2(\tau_1 \times \tau_2)$. We assume the representations: $\beta_1(s)=\sum_{k=1}^{K}\psi_{1k}(s)\eta_{1k},\ \beta_2(t)=\sum_{l=1}^{L}\psi_{2l}(t)\eta_{2l}$, and $\gamma(s,t)=\sum_{k=1}^{K}\sum_{l=1}^{L}\phi_{kl}(s,t)\nu_{k,l}$, where η_{1k}'s, η_{2l}'s, and $\nu_{k,l}$'s are the corresponding coefficients, which are unknown. Thus estimation of the parameter functions is reduced to estimation of the unknown coefficients. Using the basis function expansions we write

$$\int X_{1i}(s)\beta_1(s)ds=\sum_{k=1}^{K}\eta_{1k}\int X_{1i}(s)\psi_{1k}(s)ds \approx \sum_{k=1}^{K}\eta_{1i}a_{1k,i}$$

where $a_{1k,i} \approx \int X_{1i}(s)\psi_{1k}(s)ds$ is calculated by numerical integration techniques; see for example [24] who employ a similar technique. Similarly, we have $\int X_{2i}(t)\,\beta_2(t)dt \approx \sum_{l=1}^{L}\eta_{2l}a_{2l,i}$ and $\int X_{1i}(s)X_{2i}(t)\gamma(s,t)dsdt \approx \sum_{k=1}^{K}\sum_{l=1}^{L}\nu_{k,l}a_{k,l,i}$, where $a_{2l,i} \approx \int X_{2i}(t)\psi_{2k}(t)dt$ and $a_{k,l,i} \approx \{\int X_{1i}(s)\psi_{1k}(s)ds\}\ \{\int X_{2i}(t)\ \psi_{2k}(t)dt\}$ respectively are calculated numerically. The assumption that the functional covariates are observed on dense grids of points ensures that these integrals are approximated accurately.

To control the smoothness of the parameter functions, we take the approach [15, 43, 7, 16] of considering rich bases to model the parameter functions and adding a "roughness" penalty to the least squares fitting criterion. Let $\eta_1 = (\eta_{11}, \ldots, \eta_{1L})^T$; similarly define η_2 and ν. Then the parameters α, η_1, η_2 and ν are estimated by minimizing the penalized criterion:

$$\sum_{i=1}^{n}\left(Y_i-\alpha-a_{1,i}^{T}\eta_1-a_{2,i}^{T}\eta_2-a_{3,i}^{T}\nu\right)^2+P_1(\lambda_1,\eta_1)+P_2(\lambda_2,\eta_2)+P_3(\lambda_3,\lambda_4,\nu), \quad (2)$$

where $a_{1,i}$ is the K-dimensional vector of $a_{1k,i}$, $a_{2,i}$ is the L-dimensional vector of $a_{2l,i}$, and $a_{3,i}$ is the $K \times L$-dimensional vector of $a_{k,l,i}$; $P_1(\lambda_1, \eta_1)$, $P_2(\lambda_2, \eta_2)$, and $P_3(\lambda_3, \lambda_4, \nu)$ are penalty terms, and λ_1, λ_2, λ_3, λ_4 are corresponding smoothing parameters. We use penalties based on integrated p^{th} order derivatives, that is, $P_j(\lambda_j, \eta_j) = \lambda_j \|\partial^p \beta_j(s)/\partial s^p\|^2_{L^2}$, $j = 1, 2$ are the penalty terms corresponding to the main effects of the functional covariates, and $P_3(\lambda_3, \lambda_4, \nu) = \lambda_3 \|\partial^p \gamma(s,t)/\partial s^p\|^2_{L^2} + \lambda_4 \|\partial^p \gamma(s,t)/\partial t^p\|^2_{L^2}$ is the penalty corresponding to the interaction term. Here the norm $\|\cdot\|_{L^2}$ is induced by the inner product $< f, g >= \int fg$. The specification of the interaction penalty term follows from multivariate spline smoothing literature [52], and it accommodates the possibility of having different smoothness in the directions s and t. Define $\psi^{(p)}(t) = d^p \psi(t)/dt^p$ for some generic function $\psi(\cdot)$. Then it is easily seen that $P_1(\lambda_1, \eta_1) = \lambda_1 \eta_1^T P_{1p} \eta_1$, $P_2(\lambda_2, \eta_2) = \lambda_2 \eta_2^T P_{2p} \eta_2$ and $P_3(\lambda_3, \lambda_4, \nu) = \nu^T \{\lambda_3 P_{1p} \otimes I_K + \lambda_4 I_L \times P_{2p}\} \nu$, where $P_{1p} = \int \psi_1^{(p)}(s) \{\psi_1^{(p)}(s)\}^T ds$ and $P_{2p} = \int \psi_2^{(p)}(t) \{\psi_2^{(p)}(t)\}^T dt$ with $\psi_1^{(p)}(s) = (\psi_{11}^{(p)}(s), \ldots, \psi_{1K}^{(p)}(s))^T$ and $\psi_2^{(p)}(t) = (\psi_{21}^{(p)}(t), \ldots, \psi_{2L}^{(p)}(t))^T$.

Many authors have chosen to penalize integrated squared second derivatives, i.e. $p = 2$, for fitting (2); see for example [40]. In this paper, we favor penalties on the integrated squared first derivatives, i.e. $p = 1$; see also [20] who considered this idea. One major reason for this choice is that the first derivative penalty directly penalizes deviations from a non-functional model. Infinite penalties enforce constant parameters, say β_{01}, β_{02} and γ_0, as considered in the Section 1, and revert the model back to $Y_i = \alpha + \bar{X}_{1i}\beta_{01} + \bar{X}_{2i}\beta_{02} + \bar{X}_{1i}\bar{X}_{2i}\gamma_0 + \varepsilon_i$ - a standard two-way interaction model with the average of the functional variables as continuous covariates. Thus, penalizing the first derivatives shows preference for the standard interaction model's simplicity.

Using spline bases to represent the smooth effects as well as using a penalized criterion as in (2) has several advantages. First the model fitting is adapted from existing software; more about the implementation is described in Section 6. Second, additional covariate effects can be accommodated without difficulty. For example a linear effect of additional covariates as well as non-parametric effects of scalar covariates can be easily incorporated in the model using similar ideas to [28].

It is worthwhile to note that from (2) the unknown parameter functions $\beta_1(\cdot)$, $\beta_2(\cdot)$ and $\gamma(\cdot, \cdot)$ of model (1) can be identified uniquely only up to the projections onto the respective spaces that generate the X_{1i}'s, X_{2i}'s, and their tensor products. For example, the true $\beta_1(\cdot)$ may not be recovered completely; instead only its projection on the space defined by the curves $X_{1i}(\cdot)$ will be estimated. To see this, imagine a case where all $X_{1i}(\cdot)$ lie in a finite dimensional space, say $X_{1i}(s) = \sum_{\ell=1}^{q} \xi_{1i\ell} \Phi_\ell(s)$ for some orthogonal basis in $\mathscr{L}^2(\tau_1)$, $\{\Phi_\ell(\cdot)\}_\ell$. If $\beta_1(s) = \beta_1'(s) + \zeta \Psi_{q'}(s)$ such that $< \Psi_{q'}, \Phi_\ell >_{L^2} = 0$ for all $1 \le \ell \le q$, then we have $\int X_{1i}(s)\beta_1(s)ds = \int X_{1i}(s)\beta_1'(s)ds$. The situation is similar for the other two smooth effects, β_2 and γ.

The criterion in (2) has an available analytical solution. Stack the column vectors defined from (2) into individual design matrices $A_1 = [a_{11}|...|a_{1n}]^T$, $A_2 = [a_{21}|...|a_{2n}]^T$, and $A_3 = [a_{31}|...|a_{3n}]^T$. Then define an overall model design matrix $A = [1|A_1|A_2|A_3]$, and define S_λ to be a block diagonal matrix with blocks $[0, \lambda_1 P_1, \lambda_2 P_2, \lambda_3 P_1 \otimes I_L + \lambda_4 I_K \otimes P_2]$. By the standard ridge regression formula we obtain parameter estimates

$$\hat{\theta}=(\hat{\alpha},\hat{\eta}_1,\hat{\eta}_2,\nu)=(A^TA+S_\lambda)^{-1}A^TY, \quad (3)$$

and by extracting $\hat{\eta_1}$, $\hat{\eta_2}$, and $\hat{\nu}$ we obtain

$$\hat{\beta}_1(s)=\sum_{k=1}^{K}\psi_{1k}(s)\hat{\eta}_{1k};\quad \hat{\beta}_2(t)=\sum_{l=1}^{L}\psi_{2l}(t)\hat{\eta}_{2l};\quad \hat{\gamma}(s,t)=\sum_{k=1}^{K}\sum_{l=1}^{L}\phi_{kl}(s,t)\hat{\nu}_{k,l}.$$

Predicted values for the response are obtained by

$$\hat{Y}=A(A^TA+S_\lambda)^{-1}A^TY=H_\lambda Y. \quad (4)$$

Here H_λ represents the hat or influence matrix, which will be in important in Section 3.3 when discussing testing. Both prediction and estimation of the parameter functions depends on the choice of the smoothness parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4$. We discuss smoothness parameter selection in Section 2.3.

2.2. Standard Error Estimation

Estimation of confidence bands using penalized splines is a delicate issue (see [44], Chapter 6). A straightforward approach is to construct approximate point-wise errors bands on the frequentist covariance matrix $\mathrm{Cov}(\hat{\theta}) = (A^T A + S_\lambda)^{-1}A^T A(A^T A + S_\lambda)^{-1}\sigma^2$. This is the approach presented by [40] (Chapter 15) in their treatment of the simple functional linear model, and has also been used in the context of non-parametric regression [15, 24, 26]. However, we find in the simulation study of section 4 that confidence bands based on this covariance often provide point-wise under-coverage. This problem has been noticed previously for non-parametric additive models [52], and for functional linear models [24, 36]. Such under-coverage can be attributed to several important factors. First, the penalized fitting procedure provides biased estimates of θ whenever $\theta \neq 0$. Second, the fitting is conditional on the smoothing parameters whose uncertainty is not taken into account. Third, the level of bias induced by the penalty parameters can vary over the domain of the functional parameters. One possible alternative to account for bias is to use the Bayesian standard errors first proposed for smoothing splines by [51] and cubic splines in [47]. By specifying an improper prior, $f_\theta(\theta) \propto e^{-\theta^T S_\lambda \theta}$, it can be shown that $\theta|Y, \lambda \sim N(\hat{\theta}, (A^T A + S_\lambda)^{-1}\sigma^2)$ (see [52], Section 4.8). The matrix $\mathrm{Cov}_B(\hat{\theta}) = (A^T A + S_\lambda)^{-1}\sigma^2$ is known as the Bayesian covariance matrix. This matrix can be decomposed:

$$\mathrm{Cov}_B(\hat{\theta})=\begin{pmatrix} \sum_{\hat{\alpha}} & \sum_{\hat{\alpha},\widehat{\eta_1}} & \sum_{\hat{\alpha},\widehat{\eta_2}} & \sum_{\hat{\alpha},\hat{\nu}} \\ \sum_{\hat{\alpha},\widehat{\eta_1}} & \sum_{\widehat{\eta_1}} & \sum_{\widehat{\eta_1},\widehat{\eta_2}} & \sum_{\widehat{\eta_1},\hat{\nu}} \\ \sum_{\hat{\alpha},\widehat{\eta_2}} & \sum_{\widehat{\eta_1},\widehat{\eta_2}} & \sum_{\widehat{\eta_2}} & \sum_{\widehat{\eta_2},\hat{\nu}} \\ \sum_{\hat{\alpha},\hat{\nu}} & \sum_{\widehat{\eta_1},\hat{\nu}} & \sum_{\widehat{\eta_2},\hat{\nu}} & \sum_{\hat{\nu}} \end{pmatrix}=(A^TA+S_\lambda)^{-1}\sigma^2, \quad (5)$$

to obtain standard errors for the parameter estimates. The Bayesian formulation to the covariance for the estimates of $\hat{\theta}$ is important because we use it to obtain confidence intervals. For example, if we consider $\varphi(s, t) = [\varphi_1(s, t), ..., \varphi_{KL}(s, t)]$ we can obtain the covariance for interaction $\Sigma_{\hat{\gamma}(s,t)} = \varphi(s, t)^T \Sigma_{\hat{\nu}} \varphi(s, t)$. Intervals are found from $\hat{\gamma}(s, t) \sim N(E[\hat{\gamma}(s, t)], \Sigma_{\hat{\gamma}}(s, t))$ by standard linear models tools.

[51] presents Bayesian confidence intervals in a general framework that contains the functional linear models; but theoretical and numerical studies of the finite sample properties of these intervals have focused on non-parametric regression [25, 34, 38]. Broadly, these studies conclude the main issue for proper interval coverage, in the across-the-function sense, is that the bias must only represent a modest fraction of the overall mean squared error. The finite sample properties of Bayesian intervals in functional linear models is open research.

2.3. Smoothing parameter selection

There are several approaches to select the smoothing parameters λ_1, λ_2, λ_3, λ_4. One class of approaches selects the smoothing parameters to minimize a prediction error criterion, using Akaike's information criterion (AIC), cross validation or generalized cross validation (GCV); see for example [11]. A second class of approaches treats minimization of the penalized criterion as fitting an equivalent mixed effects model, where the smoothing parameters enter as variance components. The variance parameters are then estimated by maximum likelihood (ML, [2]) or restricted maximum likelihood/generalized maximum likelihood (REML/GML, [51]). It is generally known that the prediction error methods are rather unstable and may lead to occasional under-smoothing, whereas the more computationally intensive likelihood-based criteria such as REML/ML are more resistant to over-fitting and show greater numerical stability [42].

3. Extensions

3.1. Generalized Functional Interaction Models

Consider now the case when the outcome Y_i is generated from an exponential family EF(ϑ_i, ϱ) with dispersion parameter ϱ such that $E\{Y | X_{1i}(\cdot), X_{2i}(\cdot)\} = g^{-1}(\vartheta_i)$, where the linear predictor $\vartheta_i = \alpha + \int X_{1i}(s)\beta_1(s)ds + \int X_{2i}(t)\beta_2(t)dt + \int\int X_{1i}(s)X_{2i}(t)\,\gamma(s, t)dsdt$ and $g(\cdot)$ is a known link function. As in Section 2.1, decompositions using pre-determined basis functions are used for the unknown parameter functions β_1, β_2, and γ. The linear predictor can then be simplified to $\vartheta_i = \alpha + \sum_{k=1}^{K} \eta_{1k} a_{1k,i} + \sum_{l=1}^{L} \eta_{2l} a_{2l,i} + \sum_{k=1}^{K} \sum_{l=1}^{L} \nu_{k,l} a_{k,l,i}$, where K and L are chosen sufficiently large to capture the variability in the parameter functions. We then estimate the model components by minimizing (2) with the understanding that the sum of squares is now replaced by the appropriate negative log-likelihood function. For given smoothing parameters λ_1, λ_2, λ_3, and λ_4, there is an unique solution which can be obtained by a penalized version of the iteratively re-weighted least squares (see [52]). Asymptotic normality of these estimators follows from the large sample properties of maximum likelihood estimators and thus approximate confidence error bands can be determined accordingly (see for example [10]). [52] also proposes an efficient and stable methodology to select the smoothing parameters for generalized outcomes by employing a Laplace approximation to the REML/ML criteria and using a nested iteration procedure.

3.2. Noisy and Sparse Functional Predictors

Consider now the case when the functional predictors are observed on a dense grid of points, but with measurement error. In particular, instead of observing $X_1(\cdot)$ and $X_2(\cdot)$, we observe $W_{1i}(s) = X_{1i}(s) + \delta_{1i}(s)$ and $W_{2i}(t) = X_{2i}(t) + \delta_{2i}(t)$, where $\delta_{ji}(\cdot)$ for $j = 1, 2$ are white noise processes with zero-mean and constant variances σ_j^2. The methodology described in Section 2.1 can be still applicable with the difference that in the penalty criterion (2) for normal responses, or the negative likelihood analog for generalized responses, the terms $a_{1,i}$, $a_{2,i}$ and a_{3i} are calculated based on W_{1i}'s and W_{2i}'s in place of the X_{1i}'s and X_{2i}'s. One may also apply functional principal component analysis (FPCA) (discussed in [14, 50, 56]) to the noisy data and obtain the smoothed trajectories first, and then apply the estimation method on the smoothed covariates.

Consider next the situation when the proxy functional covariates are measured on sparse and/or irregular design points such that the set of all observation points is dense. A different approach is now needed as the terms $a_{1,i}$, $a_{2,i}$ and a_{3i} cannot be estimated accurately any longer by usual numerical integration methods. Instead, we estimate the trajectories of the underlying functional predictors X_{1i}, X_{2i} first by using FPCA, and then the approach outlined in Section 2.1 can be readily applied.

3.3. Hypothesis Testing

An advantage of our fitting approach is that it facilitates hypothesis testing based on the Wald-type test of [53]. The test applies to any exponential family response, and produces p-values directly from the software implementation described in section 6. This test could be especially useful as a model selection tool in functional linear models. We explain this next for testing the null hypothesis that there is no interaction.

Consider testing the hypothesis

$$H_0{:}\gamma(s,t)=0\ \forall s,t \quad \text{vs. } H_A{:}\gamma(s,t)\neq 0 \quad \text{for some } s,t. \qquad (6)$$

The intuition for testing is as follows. Define $\mu_\gamma = [\mu_{11}, \ldots, \mu_{1n}]^T$ be a vector of signals that correspond to interaction for each subject; where $\mu_{\gamma i} = \int\int X_{1i}(s)X_{2i}(t)\gamma(s, t)dsdt$ for $i = 1, \ldots, n$. Since the null hypothesis implies $\mu_\gamma \equiv 0$, we can base the test procedure off $\hat{\mu_\gamma}$.

From the proposed fitting procedure in (2) $\mu_{\gamma i}=a_{3i}^T\nu$, and therefore $\hat{\mu}_{\gamma i}=a_{3i}^T\hat{\nu}$. It follows that $\hat{\mu_\gamma} = A_3\hat{\nu}$ where $A_3 = [a_{31}|\ldots|a_{3n}]^T$. If the response is normally distributed, from the Bayesian covariance matrix $\Sigma_{\hat{\nu}}$ described in Section 2.2, and linear models tools

$$\hat{\mu}_\gamma \sim N(E(\hat{\mu}), \textstyle\sum_{\hat{\mu}_{\hat{\gamma}}}), \qquad (7)$$

for $E(\hat{\mu}) = A_3E(\hat{\nu})$ and $\sum_{\hat{\mu}_{\hat{\gamma}}} = A_3 \sum_{\hat{\nu}} A_3^T$. For responses generated from any exponential family the normality of $\hat{\mu}$ is valid asymptotically. The test statistic is based off the quadratic form

$$T_r = \hat{\mu}_\gamma^T \sum_{\hat{\mu}_\gamma}^{r-} \hat{\mu}_\gamma,$$

where $\sum_{\hat{\mu}_\gamma}^{r-}$ is a generalized rank-r pseudo-inverse of $\Sigma_{\hat{\mu}_\gamma}$ defined by [53]. Here r corresponds to the effective degrees of freedom as defined by the trace of the lower diagonal KL elements of $2H_\lambda - H_\lambda H_\lambda$, where H_λ is the hat matrix from (4). If r is an integer, under the null hypothesis T_r follows an asymptotic χ_r^2 distribution. When r is non-integer the asymptotic null distribution of T_r is non-standard, and p-values are calculated according to [53].

The key assumption in testing for interaction is that the Bayesian covariance matrix $\Sigma_{\hat{\nu}}$ accounts for the added uncertainty due to the bias in the estimated coefficient parameters. One way to assess this is through point-wise confidence interval coverage. In our simulation we observe the confidence intervals for the functional parameters produced by the Bayesian standard errors often provide over-coverage, which is evidence toward the testing procedure being valid.

4. Simulation

In this section we perform a numerical study of our method. The primary objective of this simulation is to evaluate our procedure, in terms of both parameter estimation and predictive performance. The functional parameter estimates are evaluated in terms of the 1) bias, 2) consistency, and 3) confidence interval coverage. Prediction is assessed in terms of estimates of the residual variance for gaussian data and mis-classification rates for bernoulli data. A secondary objective of this study is to investigate the effects of model mis-specification. The results show that fitting a purely additive model when interaction is present may lead to biased estimates but fitting our approach when the true model is in fact additive does not result in significant loss of accuracy in estimation.

4.1. Design and Assessment

The functional covariates $X_{ji}(s) = \phi_j^T(s)\xi_{ji}$, $j = 1, 2$, are generated so that $\xi_{1i} \sim MVN(0, \Sigma)$ and $\xi_{2i} \sim MVN(0, \Sigma)$ with $\Sigma = \text{diag}(8, 4, 4, 2, 2, 1, 1)$, and $\varphi_1(s) = [1, \sin(\pi s), \cos(\pi s), \sin(3\pi s), \cos(3\pi s), \sin(4\pi s), \cos(4\pi s)]$ and $\varphi_2(t) = [1, \sin(\pi t), \cos(\pi t), \sin(2\pi t), \cos(2\pi t), \sin(4\pi t), \cos(4\pi t)]$. We generate the observed functional covariates both with and without independent measurement error, according to the model $W_{1i}(s) = X_{1i}(s) + \delta_{1i}(s)$ and $W_{2i}(t) = X_{2i}(t) + \delta_{2i}(t)$, such that for $j = 1, 2$, δ_{ji} is a white noise process with $\sigma_\delta^2 = 0$, 1/4, or 4. For the parameter functions, the main effects are defined as $\beta1(s) = 2\cos(3\pi s)$, a truly functional signal, and $\beta_2(t) = 0.5$, constant and non-dependent on t. We consider two interaction parameters: $\gamma_1(s, t) = 0$, corresponding to an additive model, and $\gamma_2(s, t) = \sin(\pi s)\sin(\pi t)$, a non-trivial interaction.

All functions are evaluated at $H = 100$ equally spaced points over $s, t \in [0, 1]$. We used Riemann sums to approximate $\mu_{ji} = \int X_{ji}(s)\beta_j(s)ds$, $j = 1, 2$, and $\mu_{3i} = \int X_{1i}(s)X_{2i}(t)\gamma(s, t)dsdt$. We consider two cases: (*A*) $Y_i \sim N(\alpha + \mu_{1i} + \mu_{2i} + \mu_{3i}, 1)$ and (*B*) $Y_i \sim Bern\{(e^{\alpha+\mu_{1i}+\mu_{2i}+\mu_{3i}})/(1 + e^{\alpha+\mu_{1i}+\mu_{2i}+\mu_{3i}})\}$. We use sample sizes $n = 100$, 200, and 500 for (A); and $n = 300$ and 500 for (B). For each generated sample, we observe $\{Y_i, W_{1i}(s), W_{2i}(t)\}_{i=1}^{n}$. In all our simulations, we chose $\Psi_1(s)$ and $\Psi_2(t)$ to be cubic B-spline basis functions with 10 equally spaced internal knots, and penalize integrated squared first derivatives. The penalty parameters were estimated using REML, or with the Laplace approximation to REML for Gaussian and Bernoulli data, respectively. For comparison purposes, we also fit the additive functional linear model with the same model specifications for bases, penalty, and roughness penalty selection procedure.

We ran 1000 Monte Carlo simulations for each setting described above. Performance was assessed on the aggregate over all Monte Carlo runs, and the entire grids $s, t \in [0, 1]$, for each functional parameter. We evaluated estimates in terms of mean integrated squared error: $MISE(\hat{\beta}_1)=\sum_{j=1}^{1000}\sum_{h=1}^{H}\{\hat{\beta}_{1j}(s_h)-\beta_1(s_h)\}^2/(1000 \cdot H)$, where $\hat{\beta_{1j}}$ is the estimated parameter for the j^{th} simulated dataset. Also reported are mean point-wise $(1-\alpha)100\%$ confidence interval coverages:

$MCI(\hat{\beta}_1)=\sum_{i=1}^{1000}\sum_{h=1}^{H} I[\beta_{1i}(s_h) \in \{\hat{\beta}_{1i}(s_h) \pm z_{\alpha/2}\widehat{SE}(\hat{\beta}_{1i}(s_h))\}]/(1000 \cdot H)$. Predictive performance for the Gaussian data is evaluated by average prediction error (APE):

$\text{APE}=\sum_{j=1}^{1000}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2/(1000 \cdot n)$. The optimal APE equals the residual variance of 1, APEs below 1 indicate over-fitting of the model to the data, and APEs above 1 suggest under-fitting of the model. For the Bernoulli data we focus on the mis-classification (MC) rate: $\text{MC}=\sum_{j=1}^{1000}\sum_{i=1}^{n} I(y_i=\hat{y}_i)/(1000 \cdot n)$, where $\hat{y}_i = 0$ if $\hat{\pi}_i \leq .5$ and $\hat{y}_i = 1$ otherwise.

4.2. Results

Focus first on the results without measurement error in Table 1. For the situation where Gaussian data is generated with the interaction term γ_2 (non-trivial interaction effect), and the interaction model is correctly used, the parameter function estimates have monotonically decreasing MISEs with increasing sample size. The APEs are all below 1 which suggests over-fitting on the average, however this over-fitting is only moderate and decreases with sample size. In contrast, when the additive model is incorrectly used, the estimates are affected adversely for all metrics of evaluation. There is a marked increase in the MISEs for estimation of β_1 and β_2, and a large loss of prediction power even for increasing sample size.

We compare these results of mis-specification to the situation where data is generated with γ_1 (an additive model). At sample size $n = 100$, fitting an interaction model resulted in moderately increased MISEs and lower APEs, due to more over-fitting. Nevertheless, application of the additive and interaction model gave highly similar results for sample sizes of 200 and 500. The key is that with sufficient sample size to empower selection of the smoothing parameters, the model chooses the additive fit on it's own.

The frequentist intervals tend to provide under-coverage, while the Bayesian intervals tend to give over-coverage, at the 95% nominal level. This challenging issue is not specific to the interaction model however; it persists when there is no interaction and an additive model is correctly fit. Further investigation indicates that on average, the empirical Monte Carlo standard errors of the parameter estimates are sandwiched between the average estimated frequentist and Bayesian standard errors. The over-coverage of the Bayesian intervals is a result of an over-correction for the bias caused by the penalized regression procedure.

The reduced information in the Bernoulli responses led to less efficient estimation of all parameters. One difference from the results of the Gaussian data, is that there is noticeable bias in the estimation of γ_2, and poor confidence interval coverage for interaction. However, the effects of mis-specification tell a similar story. When γ_2 is the truth and the additive model is fit, we have inflated biases, almost non-existent confidence interval coverage, and larger mis-classification rates. In contrast, if the data is generated from γ_1 and the interaction model is fit, the results are highly similar to those found when the additive model is applied.

Results for when the functional covariates are generated with measurement error appear in Appendix Tables A1 and A2. When $\sigma_\delta^2 = 1/4$ the results are highly similar to the case of no error. For $\sigma_\delta^2 = 4$ the measurement error noise is on the scale of the scores generating the true covariates, and in this case all the metrics are affected adversely.

5. AneuRisk study

To illustrate our method we focus on the AneuRisk65 data described in [46]. A broad goal of this study is to identify the relationship between the geometry of the internal carotid artery (ICA) and the presence or absence of an aneurysm downstream of the ICA. The study contains a collection of 3D angiographic images taken from 65 subjects thought to be affected by a cerebral aneurysm. Of these 65 subjects, 33 have an aneurysm located downstream of a terminal bifurcation in the ICA (upper group), 25 have an aneurysm located on the terminal bifurcation of the ICA (lower group), and 7 have no aneurysm (no-group). In this study, the presence or absence of an aneurysm downstream of the ICA is of primary of interest, and therefore the 32 subjects in the latter two groups are combined (lower group) [46]. For each subject, the images are summarized to describe the geometry of the ICA. [39] approximate the centerline of the artery in 3D space and estimate the corresponding width of the artery along this centerline in terms maximum inscribed sphere radius (MISR). [46] provide a measure of curvature of the artery in 3D space along the artery centerline. The curvature and MISR profiles observed along the ICA centerline serve as our functional predictors. In this situation, the 3D geometries of the arteries are more thoroughly described by the combination the curvature and MISR values taken along the ICA centerline, and therefore it makes sense to include a two-way interaction term in the model. Our interest is two-fold: 1) to classify the subjects using the curvature and MISR profiles with the proposed penalized spline framework; and 2) to infer whether a including a two-way interaction term between the curvature and MISR profiles helps better explain group status.

There are a few registration approaches proposed in the literature to align the profiles [54, 21, 9, 49]; for a discussion of these approaches see [45] (rejoinder). We are using the technique discussed in [49], based on the Fisher-Rao curve registration method [48]. Previous analyses with this registration approach showed similar classification results to the approaches proposed by [9] and [46].

The aligned profiles and their estimated means are shown in Figure 1; the abscissa parameter takes values from −1 to 0, where the negative values indicate the direction along the ICA opposite to the blood flow. Individuals with an aneurysm on the ICA are coded as 1, while the rest are 0. We regress this binary response on the aligned and de-meaned profiles for curvature and MISR. We apply the interaction model specified for a logistic link function, penalize the first derivative norms, and capture the effect of β_1, β_2, and γ via cubic spline bases with 5 equally spaced knots ($K = L = 7$). The number of knots are chosen to be as large as possible. The fitting procedure described later in section 6 requires the number of coefficients for model fitting to be less than sample size. Therefore, we specify $K = L = 7$ so that the penalized likelihood has $1 + 7 + 7 + 49 = 64 < 65$ coefficients. For comparison, we apply the analogous additive model to that fit in pfr, and maintain the same bases and penalization as used in the interaction model.

Figure 2 displays the estimated interaction contour. There is a significant and positive estimated effect of interaction over the region where curvature takes values from −0.5 to 0 and MISR from −0.6 to −0.2. Therefore, over these regions subjects with curvature values above the population mean, and MISRs below the population mean, should tend to be classified in the lower group. This is in line with data shown in Figure 1. Those in the lower group tend to have distinctly higher values of curvature around two sharp peaks in curvature near −0.2 and −0.3, and more often have lower values of MISR over the region of −0.6 to −0.2. The main effects estimates are shown in Figure 3. The main effects estimates are similar for the additive and interaction models. However, in the interaction model the estimate of β_1 is more slightly downward sloping the than the estimate from the additive model. Both models give positive estimates for β_2 from −1 to −0.4, and over this region the MISRs for those in the upper group tend to take values higher than for those in the lower group. For both models, the Bayesian confidence intervals for the positive MISR main effects exclude 0 in a small region around −.63.

We compare prediction in terms of the number of subjects mis-classified from the direct sample estimates using the apparent error rate (APER), and also include the leave-one-out error rate (L1ER). Observations whose estimated probability of upper group membership exceed .5 are classified as 1 and vice versa. The error rates for the additive model are 19/65 and 24/65 for the APER and L1ER respectively; and 11/65 and 22/65 for the interaction model. While the reduction in mis-classification error was less for the leave-one-out estimates, we observe that the median difference of the probability of group membership for the leave-one-out estimates still differs substantially (see Table 2 and Figure 4 below).

[46] used quadratic discriminant analysis (QDA) of the top principal component (PC) scores and achieved APER and L1ER mis-classification rates of 10/65 and 14/65. Their classification procedure is similar to ours in that QDA allows for interaction, but at the level of the PC scores. While their procedure shows better classification rates, especially for the L1ER, it is important to note that the number of principal components were chosen to minimize the L1ER criteria directly, as opposed to our automated dimension reduction with smoothing parameters selected by REML. Furthermore, a possible advantage of our model is that the parameter estimates can provide visual insight into the relation between the functional covariates and the response, while QDA is focused solely on classification. For further comparison, we estimated the main effects and interaction contour with a regression approach that uses models the top functional principal components similar to [46] (see Figures 5 and 6).

The small difference in the leave-one-out estimates from the additive and interaction model makes it difficult to determine whether including the interaction piece is helpful for this data. Therefore, we carried out a hypothesis test of the interaction effect using the procedure described in Section 3.3. The test statistic for the interaction effect $T_{7.2} = 10.1$; where $r = 7.2$ represents the reference degrees of freedom; and this led to a p-value of .19. Since this result did not show significance we also tested main effects from the additive model. For tests of $\beta_1(s) = 0$ and $\beta_2(t) = 0$, the test statistics were $T_{2.5} = 2.4$ and $T_{3.5} = 10.4$ respectively, which corresponded to p-values of .40 and .02. While only the effect of $\beta_2(t)$ was declared statistically significant, we should interpret these results with caution due to the small sample size and the fact that the testing procedure is based on asymptotics. Furthermore, separate individual FLM analyses of curvature and MISR model fitting procedure produced p-values of .02 and .01; and APERs of 23/65 and 21/65 respectively.

6. Implementation

Fitting was carried out with the gam function from the mgcv package (see [52] for details). The gam function is highly flexible and allows for the model to be fit with a variety of basis and penalty combinations. The summary output gives measures of model fit in terms of R^2 and deviance explained, automatically provides p-values for each smooth functional parameter, and allows for direct plotting of the functional parameters along with their Bayesian confidence bands. A computer code demonstrating the proposed approach using R is available at http://www4.stat.ncsu.edu/~maity/software.html.

7. Discussion

We considered a penalized spline based method for functional regression that incorporates two-way interaction effects between functional predictors. The proposed framework can handle responses from any exponential family, functional predictors measured with error or on a sparse grid, and provides hypothesis tests for individual model components. The main advantage of our framework is that it can be fit with highly flexible and readily available software, that provides detailed summaries of the model fit. These summaries can guide whether inclusion of interaction into the functional linear model is appropriate.

Mis-specification of an additive model in the face of interaction has adverse effects. Through simulation we found that failure to account for interaction led to poor parameter estimation, diminished confidence interval coverage, and lost prediction power. In contrast, mis-specification of the interaction model showed negligible adverse effects, especially for moderate or large sample sizes. Confidence interval coverage was an issue in the simulation study, but was not specific to the interaction model. Evaluation of Bayesian standard errors have mostly focused on non-parametric regressions and require further investigation for functional linear models. This issue is especially important because of the correspondence between the Bayesian covariance matrix and the proposed hypothesis testing procedure in section 3.3. Evaluation of this hypothesis testing procedure is part of our future research.

There are several other possible directions for future work. One main direction that we currently investigate is the development of alternative hypothesis tests for the interaction effect with greater power in finite samples. Equally important would be the theoretical study of the asymptotic distributions of the parameter estimators, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\gamma}$, akin to that provided by [37] in the situation of an additive model. Our paper provides a simple approach to account for interaction in a linear fashion; extensions to more flexible non-parametric dependence is part of our future research. Finally, the effect of dependence in the functional covariates will be rigorously investigated.

Acknowledgments

This research was partially supported by grant numbers DMS 1007466, DMS 0454942, 1R01NS085211-01 and 2R01MH086633-05A1 (Staicu) and R00ES 017744 (Maity and Usset). The content is solely the responsibility of the authors and does not necessarily represent the social views of the National Institutes of Health.

References

1. Amato U, Antoniadis A, De Feis I. Dimension reduction in functional regression with applications. Computational Statistics & Data Analysis. 2006; 50(9):2422–2446.
2. Anderssen R, Bloomfield P. A time series approach to numerical differentiation. Technometrics. 1974; 16(1):69–75.
3. Aneiros-Pérez G, Vieu P. Semi-functional partial linear regression. Statistics & Probability Letters. 2006; 76(11):1102–1110.
4. Aneiros-Pérez G, Vieu P. Nonparametric time series prediction: A semi-functional partial linear modeling. Journal of Multivariate Analysis. 2008; 99(5):834–857.
5. Aneiros-Pérez G, Vieu P. Partial linear modelling with multi-functional covariates. Computational Statistics. 2015:1–25.
6. Bongiorno, EG.; Salinelli, E.; Goia, A.; Vieu, P. Contributions in infinite-dimensional statistics and related topics. Società Editrice Esculapio; 2014.
7. Cardot H, Ferraty F, Sarda P. Spline estimators for the functional linear model. Statistica Sinica. 2003; 13:571–591.
8. Chen D, Hall P, Müller HG. Single and multiple index functional regression models with nonparametric link. The Annals of Statistics. 2011; 39(3):1720–1747.
9. Cheng W, Dryden IL, Hitchcock DB, Le H, et al. Analysis of aneurisk65 data: Internal carotid artery shape analysis. Electronic Journal of Statistics. 2014; 8(2):1905–1913.
10. Cox, DR.; Hinkley, DV. Theoretical statistics. CRC Press; 1979.
11. Craven P, Wahba G. Smoothing noisy data with spline functions. Numerische Mathematik. 1978; 31(4):377–403.
12. de Boor, C. A practical guide to splines. Vol. 27. Springer-Verlag; New York: 1978.

13. Delsol L. No effect tests in regression on functional variable and some applications to spectrometric studies. Computational Statistics. 2013; 28(4):1775–1811.
14. Di CZ, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel functional principal component analysis. Annals of Applied Statistics. 2009; 3(1):458–488. [PubMed: 20221415]
15. Eilers PH, Marx BD. Flexible smoothing with b-splines and penalties. Statistical science. 1996; 11(2):89–102.
16. Eilers PH, Marx BD. Multidimensional penalized regression signal regression. Technometrics. 2005; 47(1):13–22.
17. Ferraty F, Goia A, Salinelli E, Vieu P. Functional projection pursuit regression. Test. 2013; 22(2): 293–320.
18. Ferraty, F.; Vieu, P. Nonparametric functional data analysis: theory and practice. Springer Science & Business Media; 2006.
19. Ferraty F, Vieu P. Additive prediction and boosting for functional data. Computational Statistics & Data Analysis. 2009; 53(4):1400–1413.
20. Gertheiss J, Maity A, Staicu AM. Variable selection in generalized functional linear models. Stat. 2013; 2(1):86–101. [PubMed: 25132690]
21. Gervini D, et al. Analysis of aneurisk65 data: Warped logistic discrimination. Electronic Journal of Statistics. 2014; 8(2):1930–1936.
22. Goia A. A functional linear model for time series prediction with exogenous variables. Statistics & Probability Letters. 2012; 82(5):1005–1011.
23. Goia A, Vieu P. A partitioned single functional index model. Computational Statistics. 2013:1–20.
24. Goldsmith J, Bobb J, Crainiceanu C, Caffo B, Reich R. Penalized functional regression. Journal of Computational and Graphical Statistics. 2011; 20(4):830–851. [PubMed: 22368438]
25. Gu C, Wahba G. Journal of Computational and Graphical Statistics. 1993; 2(1):97–117.
26. Hastie, TJ.; Tibshirani, RJ. Generalized Additive Models. Vol. 43. CRC Press; 1990.
27. Horváth, L.; Kokoszka, P. Inference for functional data with applications. Vol. 200. Springer Science & Business Media; 2012.
28. Ivanescu AE, Staicu A-M, Scheipl F, Greven S. Penalized function-on-function regression. Computational Statistics. 2014:1–30.
29. James G. Generalized linear models with functional predictors. Journal of the Royal Statistical Society Series B. 2002; 64(3):411–432.
30. James GM, Silverman BW. Functional adaptive model estimation. Journal of the American Statistical Association. 2005; 100(470):565–576.
31. Kudraszow NL, Vieu P. Uniform consistency of knn regressors for functional variables. Statistics & Probability Letters. 2013; 83(8):1863–1870.
32. Lian H. Functional partial linear model. Journal of Nonparametric Statistics. 2011; 23(1):115–128.
33. Maity A, Huang JZ. Partially linear varying coefficient models stratified by a functional covariate. Statistics & probability letters. 2012; 82(10):1807–1814. [PubMed: 22904586]
34. Marra G, Wood SN. Coverage properties of confidence intervals for generalized additive model components. Scandinavian Journal of Statistics. 2012; 39(1):53–74.
35. McLean MW, Hooker G, Ruppert D. Restricted likelihood ratio tests for linearity in scalar-on-function regression. 2015 Pre-print, arXiv:1310:5811v1.
36. McLean MW, Hooker G, Staicu AM, Scheipl F, Ruppert D. Functional generalized additive models. Journal of Computational and Graphical Statistics. 2014; 23(1):249–269. [PubMed: 24729671]
37. Muller HG, Stadtmuller U. Generalized functional linear models. The Annals of Statistics. 2005; 33(2):774–805.
38. Nychka D. Bayesian confidence intervals for smoothing splines. Journal of the American Statistical Association. 1988; 83(404):1134–1143.
39. Piccinelli M, Bacigaluppiz S, Boccardi E, Ene-Iordache B. Influence of internal carotid artery geometry on aneurysm location and orientation: a computational geometry study. Neurosurgery. 2011; 68(5):1270–1285. [PubMed: 21273931]

40. Ramsay, J.; Silverman, BW. Functional data analysis. Wiley Online Library; 2005.
41. Ramsay JO, Dalzell C. Some tools for functional data analysis. Journal of the Royal Statistical Society. Series B (Methodological). 1991:539–572.
42. Reiss PT, Ogden TR. Smoothing parameter selection for a class of semiparametric linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2009; 71(2):505–523.
43. Ruppert D. Selecting the number of knots for penalized splines. Journal of Computational and Graphical Statistics. 2002; 11(4):735–757.
44. Ruppert, D.; Wand, MP.; Carroll, RJ. Semiparametric regression. Vol. 12. Cambridge University Press; 2003.
45. Sangalli LM, Secchi P, Vantini S. Rejoinder: Analysis of aneurisk65 data. Electronic Journal of Statistics. 2014; 8(2):1937–1939.
46. Sangalli LM, Secchi P, Vantini S, Veneziani A. A case study in exploratory functional data analysis: geometrical features of the internal carotid artery. Journal of the American Statistical Association. 2009; 104(485)
47. Silverman BW, et al. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. Journal of the Royal Statistical Society, Series B. 1985; 47(1):1–52.
48. Srivastava A, Klassen E, Joshi SH, Jermyn IH. Shape analysis of elastic curves in euclidean spaces. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2011; 33(7):1415–1428.
49. Staicu AM, Lu X, et al. Analysis of aneurisk65 data: Classification and curve registration. Electronic Journal of Statistics. 2014; 8(2):1914–1919.
50. Staniswalis JG, Lee JJ. Nonparametric regression analysis of longitudinal data. Journal of the American Statistical Association. 1998; 93(444):1403–1418.
51. Wahba, G. Spline models for observational data. Vol. 59. Siam; 1990.
52. Wood, SN. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC; Boca Raton, FL: 2006.
53. Wood SN. On p-values for smooth components of an extended generalized additive model. Biometrika. 2013; 100(1):221–228.
54. Xie Q, Kurtek S, Srivastava A, et al. Analysis of aneurisk65 data: Elastic shape registration of curves. Electronic Journal of Statistics. 2014; 8(2):1920–1929.
55. Yang W-H, Wikle CK, Holan SH, Wildhaber ML. Ecological prediction with nonlinear multivariate time-frequency functional data models. Journal of Agricultural, Biological, and Environmental Statistics. 2013:1–25.
56. Yao F, Müller HG, Clifford AJ, Dueker SR, Follett J, Lin Y, Buchholz BA, Vogel JS. Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. Biometrics. 2003; 59(3):676–685. [PubMed: 14601769]
57. Zhao N, Bell DA, Maity A, Staicu AM, Joubert BR, London SJ, Wu MC. Global analysis of methylation profiles from high resolution cpg data. Genetic epidemiology. 2015; 39(2):53–64. [PubMed: 25537884]
58. Zhou J, Chen M. Spline estimators for semi-functional linear model. Statistics & Probability Letters. 2012; 82(3):505–513.

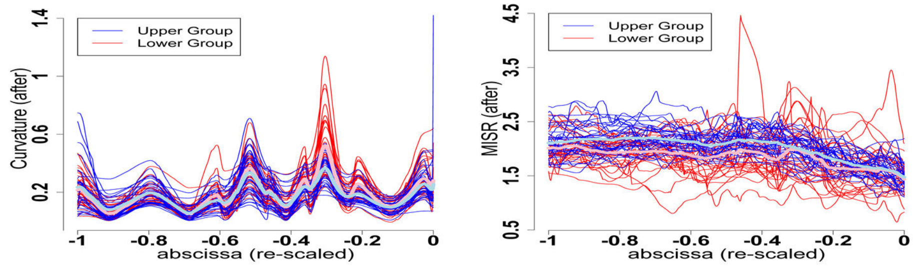

Figure 1.

Aligned curvature (left) and MISR (right) functions obtained from Fisher Rao curve registration. Color indicates group membership: blue for individuals with an aneurysm present on the ICA (upper group) and red for individuals where the aneurysm in absent on the ICA (lower-group). The thicker light blue and pink lines represent the group means for the upper and lower groups respectively.

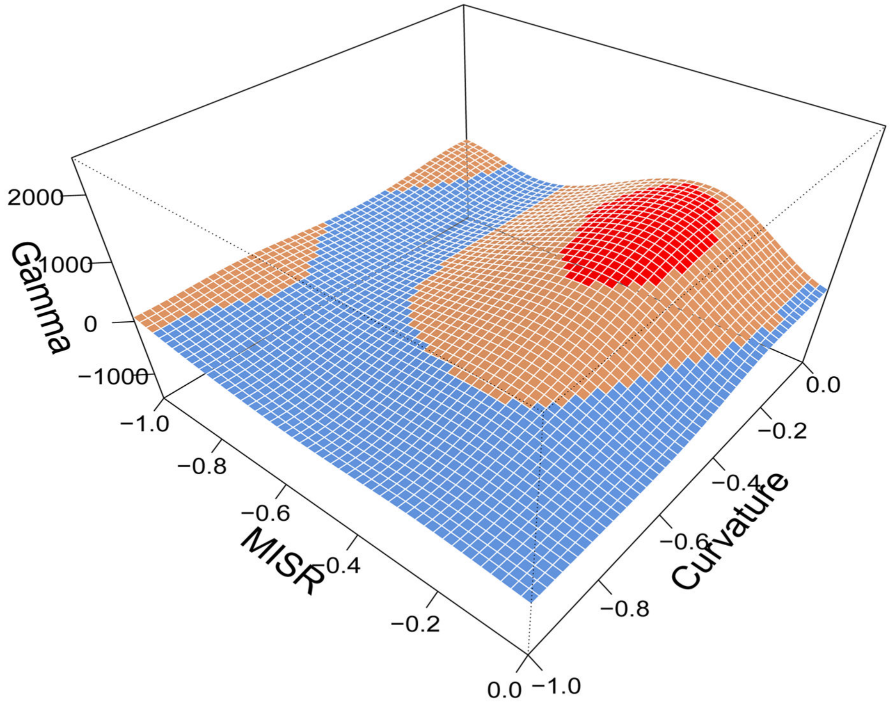

Figure 2.
The estimated interaction contour along with measures of significance. Color-coding: dark red/blue is for positive/negative significant values (at 95% level), while light red/blue is used for positive/negative values.

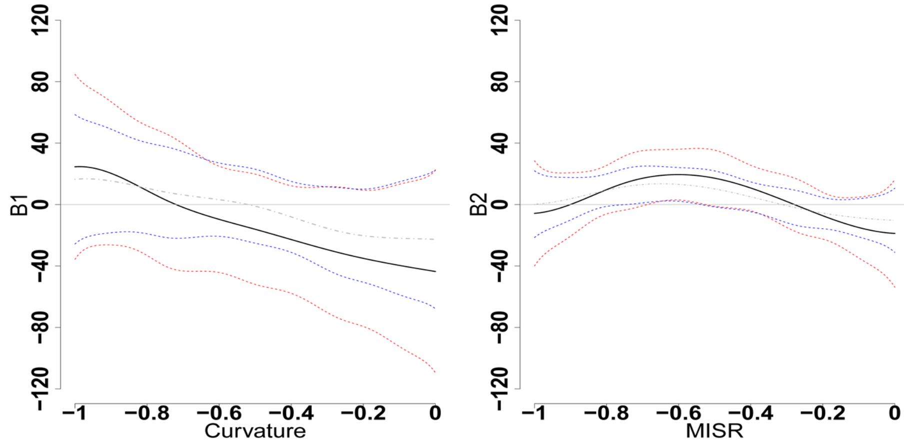

Figure 3.
Results for the AneuRisk study using our proposed penalized splines methodology. The left and right plot show the main effects (black solid line) and point-wise 95% Bayesian confidence bands (red dashed) using the functional interaction model; overlaid are the estimated main effects using the functional additive model and the corresponding point-wise 95% Bayesian confidence bands (blue dotted).

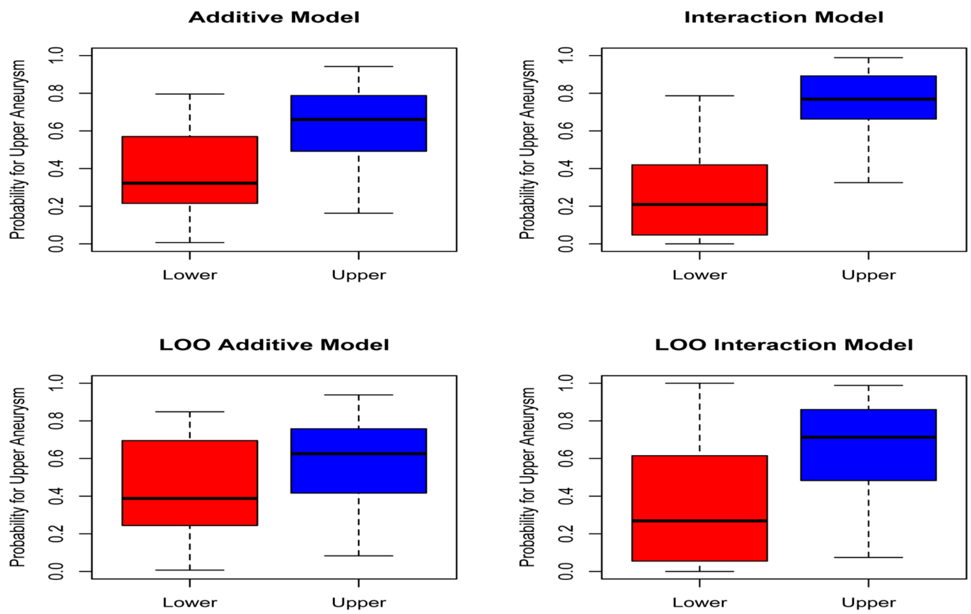

Figure 4.

The top row gives the probability estimates of an aneurysm on the ICA from the additive (left) and interaction (right) model. The bottom row corresponds to the leave-one-out (LOO) estimates from the spline-based additive (left) and interaction (right) model.

Author Manuscript

Author Manuscript

Author Manuscript

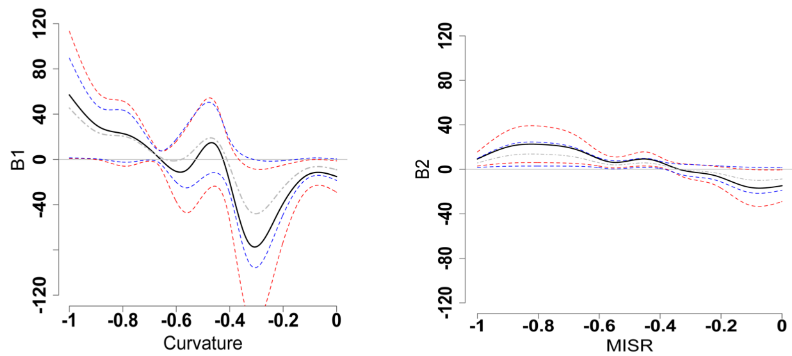

Figure 5.

Results for the AneuRisk study data using a functional principal components regression (FPCR) that incorporates interaction. The left and right plot show the main effects (black solid line) and point-wise 95% confidence bands (red dashed) using the functional interaction model estimated with FPCR; overlaid are the estimated main effects using the FPCR functional additive model and the corresponding point-wise 95% confidence bands (blue dotted).

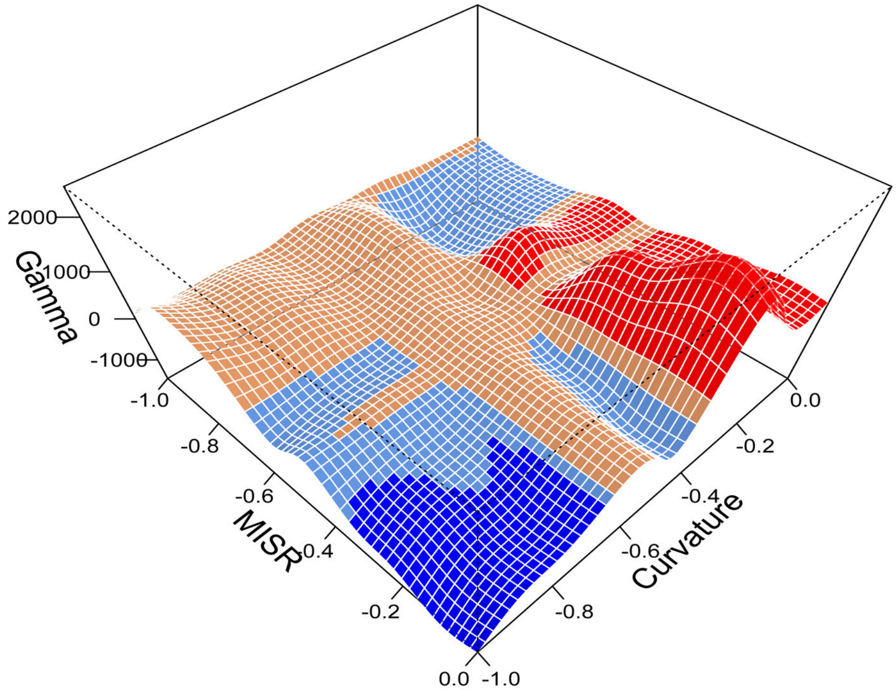

Figure 6.

Results for the AneuRisk study data using a functional principal components regression (FPCR) that incorporates interaction. This plot displays the estimated interaction contour from FPCR along with measures of significance. Color-coding: dark red/blue is for positive/negative significant values (at 95% level), while light red/blue is used for positive/negative values.

Table 1

Simulation results when the functional covariates are observed without error (top) and with measurement error (bottom). The results represent 100 times the mean integrated squared biases (ISB), mean integrated square errors (MISE), mean confidence interval coverages corresponding to the frequentist (MCI_F) and Bayesian standard errors (MCI_B), averaged prediction errors (APE) for the continuous responses, and mis-classification rates (MC) for the Bernoulli data, over 1000 runs for β_1, β_2, and γ, when the true model (True) is additive (Add) or involving non-trivial interaction effect (Int) and fit with model specified in the column 'Fit'. The standard errors for MISEs are in parentheses, while standard errors for all other metrics were less than 1.

$\sigma_\delta^2=0$			β_1				β_2				γ				
	True	**Fit**	**ISB**	**MISE**	MCI_F	MCI_B	**ISB**	**MISE**	MCI_F	MCI_B	**ISB**	**MISE**	MCI_F	MCI_B	**APE**
Gaussian	Add	Add	0.2	10.2 (0.2)	93.9	100.0	0.0	0.4 (0.0)	90.7	95.1	-	-	-	-	91.9
n = 100		Int	0.1	18.5 (1.1)	83.3	99.4	0.0	1.5 (0.1)	76.7	89.5	0.0	1.1 (0.4)	73.7	94.9	83.2
	Int	Add	21.3	88.3 (2.0)	73.7	84.2	0.0	10.4 (0.6)	74.8	81.2	-	-	-	-	1689.3
		Int	0.2	20.1 (1.0)	81.6	99.2	0.0	6.2 (4.3)	74.4	89.0	0.2	3.9 (0.1)	89.7	99.0	73.2
	Add	Add	0.1	7.3 (0.2)	95.0	100.0	0.0	0.2 (0.0)	92.7	96.6	-	-	-	-	95.6
n = 200		Int	0.1	7.3 (0.2)	95.0	100.0	0.0	0.2 (0.0)	92.3	96.5	0.0	0.1 (0.0)	87.8	96.4	93.9
	Int	Add	4.4	40.3 (0.9)	89.1	98.1	0.0	5.3 (0.3)	73.9	80.4	-	-	-	-	1741.1
		Int	0.0	7.0 (0.2)	94.8	100.0	0.0	0.2 (0.0)	91.3	96.3	0.3	1.4 (0.0)	89.7	99.9	87.9
	Add	Add	0.0	5.8 (0.2)	94.9	100.0	0.0	0.1 (0.0)	93.0	96.9	-	-	-	-	98.6
n = 500		Int	0.0	5.8 (0.2)	94.9	100.0	0.0	0.1 (0.0)	92.7	96.8	0.0	0.0 (0.0)	88.2	96.6	97.9
	Int	Add	0.9	20.7 (0.4)	92.7	99.8	0.0	1.8 (0.1)	77.1	82.5	-	-	-	-	1814.1
		Int	0.0	5.8 (0.2)	94.9	100.0	0.0	0.1 (0.0)	92.5	96.6	0.2	0.9 (0.0)	92.2	100.0	95.6
	True	Fit	ISB	MISE	MCI_F	MCI_B	ISB	MISE	MCI_F	MCI_B	ISB	MISE	MCI_F	MCI_B	MC
Logistic	Add	Add	0.4	18.1 (0.5)	93.7	99.9	0.0	1.2 (0.1)	93.1	96.4	-	-	-	-	27.9
n = 300		Int	0.3	18.7 (0.5)	93.7	99.9	0.0	1.4 (0.1)	93.2	97.0	0.0	0.3 (0.0)	89.9	97.0	27.6
	Int	Add	59.9	67.4 (0.7)	32.7	60.4	12.2	12.9 (0.1)	3.1	4.9	-	-	-	-	41.6
		Int	0.7	24.5 (0.6)	92.5	99.7	0.0	2.3 (0.2)	90.1	94.7	2.6	6.2 (0.1)	64.3	82.2	20.5
	Add	Add	0.2	13.2 (0.3)	94.2	99.9	0.0	0.7 (0.0)	92.5	95.9	-	-	-	-	28.1
n = 500		Int	0.2	13.4 (0.3)	94.3	99.9	0.0	0.7 (0.1)	92.5	96.3	0.0	0.2 (0.0)	89.2	96.7	27.9

$\sigma_\delta^2=0$			β_1				β_2				γ				
	True	**Fit**	**ISB**	**MISE**	***MCI_F***	***MCI_B***	**ISB**	**MISE**	***MCI_F***	***MCI_B***	**ISB**	**MISE**	***MCI_F***	***MCI_B***	**APE**
	Int	Add	56.8	61.9 (0.5)	25.9	55.4	12.1	12.5 (0.1)	2.0	3.6	-	-	-	-	41.8
		Int	0.3	16.4 (0.4)	93.8	99.9	0.0	1.2 (0.1)	91.2	94.6	1.7	4.5 (0.1)	72.4	89.5	20.7

Table 2

Confusion matrices for additive model (left) and interaction model (right).

	Additive Model				Interaction Model			
	APER (19/65)		L1ER (24/65)		APER (11/65)		L1ER (22/65)	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Lower	22	10	21	11	26	6	19	13
Upper	9	24	13	20	5	28	9	24

Table A1

Simulation results when the functional covariates are observed without error. The results represent 100 times the mean integrated squared biases (ISB), mean integrated square errors (MISE), mean confidence interval coverages corresponding to the frequentist (MCI_F) and Bayesian standard errors (MCI_B), averaged prediction errors (APE) for the continuous responses, and mis-classification rates (MC) for the Bernoulli data, over 1000 runs for β_1, β_2, and γ, when the true model (True) is additive (Add) or involving non-trivial interaction effect (Int) and fit with model specified in the column 'Fit'. The standard errors for the mean MISEs are in parentheses, while standard errors for all other metrics were less than 1.

$\sigma_\delta^2 = 1/4$			β_1				β_2				γ				
	True	**Fit**	**ISB**	**MISE**	MCI_F	MCI_B	**ISB**	**MISE**	MCI_F	MCI_B	**ISB**	**MISE**	MCI_F	MCI_B	**APE**
Gaussian	Add	Add	0.1	9.7 (0.2)	94.2	99.0	0.0	0.3 (0.0)	92.3	96.4	-	-	-	-	92.3
n = 100		Int	0.1	15.5 (0.6)	83.4	99.6	0.0	1.4 (0.1)	77.0	88.8	0.0	0.6 (0.1)	73.3	94.3	83.7
	Int	Add	20.8	89.5 (2.2)	74.1	84.5	0.0	11.4 (0.8)	73.1	79.9	-	-	-	-	1734.5
		Int	0.1	17.1 (0.8)	82.0	99.5	0.0	1.8 (0.2)	74.8	88.5	0.2	3.8 (0.1)	78.7	99.1	73.1
	Add	Add	0.0	6.8 (0.2)	94.5	100.0	0.0	0.2 (0.0)	92.2	95.7	-	-	-	-	95.6
n = 200		Int	0.0	6.9 (0.2)	94.4	100.0	0.0	0.2 (0.0)	91.6	95.4	0.0	0.0 (0.0)	88.9	96.7	94.2
	Int	Add	4.5	43.2 (1.2)	88.6	97.6	0.0	4.9 (0.3)	75.6	81.8	-	-	-	-	1771.9
		Int	0.0	7.0 (0.2)	94.5	100.0	0.0	0.2 (0.0)	91.7	96.3	0.3	1.4 (0.0)	90.1	99.9	87.9
	Add	Add	0.0	4.4 (0.1)	95.1	100.0	0.0	0.1 (0.0)	91.9	96.2	-	-	-	-	97.7
n = 500		Int	0.0	4.4 (0.1)	95.0	100.0	0.0	0.1 (0.0)	91.6	96.2	0.0	0.0 (0.0)	89.0	97.6	97.5
	Int	Add	1.0	21.6 (0.5)	92.1	99.7	0.0	2.3 (0.2)	73.5	80.3	-	-	-	-	1834.3
		Int	0.0	4.5 (0.1)	95.0	100.0	0.0	0.1 (0.0)	91.4	95.9	0.2	0.9 (0.0)	92.4	100.0	93.8
	True	Fit	ISB	MISE	MCI_F	MCI_B	ISB	MISE	MCI_F	MCI_B	ISB	MISE	MCI_F	MCI_B	MC
Logistic	Add	Add	0.2	17.6 (0.4)	93.9	99.8	0.0	1.2 (0.2)	92.3	96.5	-	-	-	-	27.9
n = 300		Int	0.1	18.2 (0.5)	93.9	99.8	0.0	1.4 (0.2)	92.6	96.5	0.0	0.3 (0.0)	88.4	96.9	27.5
	Int	Add	58.0	65.9 (0.7)	32.7	60.4	11.8	12.5 (0.1)	3.1	12.9	-	-	-	-	41.5
		Int	0.5	23.9 (0.7)	92.8	99.7	0.0	2.2 (0.2)	90.8	94.7	2.4	6.4 (0.1)	64.8	82.2	20.0
	Add	Add	0.2	12.9 (0.3)	94.2	100.0	0.0	0.7 (0.1)	92.4	96.0	-	-	-	-	27.9
n = 500		Int	0.1	13.1 (0.3)	94.2	100.0	0.0	0.8 (0.1)	91.9	95.9	0.0	0.2 (0.0)	89.1	97.8	27.7
	Int	Add	55.6	60.9 (0.5)	26.3	57.2	12.0	12.5 (0.1)	1.8	3.2	-	-	-	-	41.7

$\sigma_\delta^2 = 1/4$		β_1				β_2				γ				
True	**Fit**	**ISB**	**MISE**	***MCI$_F$***	***MCI$_B$***	**ISB**	**MISE**	***MCI$_F$***	***MCI$_B$***	**ISB**	**MISE**	***MCI$_F$***	***MCI$_B$***	**APE**
	Int	0.3	16.4 (0.4)	93.6	99.9	0.0	1.2 (0.1)	91.2	94.6	1.7	4.5 (0.1)	71.9	89.5	20.7

Table A2

Simulation results when the functional covariates are observed with measurement error ($\sigma_\delta^2=4$). The results represent 100 times the mean integrated squared biases (ISB), mean integrated square errors (MISE), mean confidence interval coverages corresponding to the frequentist (MCI_F) and Bayesian standard errors (MCI_B), averaged prediction errors (APE) for the continuous responses, and mis-classification rates (MC) for the Bernoulli data, over 1000 runs for β_1, β_2, and γ, when the true model (True) is additive (Add) or involving non-trivial interaction effect (Int) and fit with model specified in the column 'Fit'. The standard errors for the mean MISEs are in parentheses, while standard errors for all other metrics were less than 1.

$\sigma_\delta^2=4$			β_1				β_2				γ				
	True	**Fit**	**ISB**	**MISE**	MCI_F	MCI_B	**ISB**	**MISE**	MCI_F	MCI_B	**ISB**	**MISE**	MCI_F	MCI_B	**APE**
Gaussian	Add	Add	5.5	18.8 (0.3)	-	96.6	0.0	0.4 (0.0)	-	95.9	-	-	-	-	117.9
n = 100		Int	4.9	29.0 (0.7)	-	86.2	0.0	4.7 (0.9)	-	86.3	0.0	13.1 (3.7)	-	72.3	99.0
	Int	Add	38.9	98.2 (2.1)	-	78.2	0.0	10.4 (0.6)	-	81.6	-	-	-	-	1715.6
		Int	5.1	39.7 (1.1)	-	85.7	0.0	7.7 (1.2)	-	84.4	0.7	35.3 (7.2)	-	76.1	117.0
	Add	Add	4.8	13.2 (0.2)	-	95.6	0.0	0.2 (0.0)	-	95.1	-	-	-	-	124.1
n = 200		Int	4.8	13.2 (0.2)	-	95.5	0.0	0.2 (0.0)	-	94.7	0.0	0.1 (0.0)	-	88.5	121.8
	Int	Add	14.4	50.7 (1.1)	-	94.2	0.0	5.1 (0.3)	-	80.7	-	-	-	-	1764.2
		Int	5.3	15.8 (0.2)	-	96.6	0.0	0.4 (0.0)	-	92.2	1.1	2.6 (0.0)	-	76.6	165.7
	Add	Add	4.4	9.5 (0.1)	-	91.0	0.0	0.1 (0.0)	-	92.8	-	-	-	-	128.7
n = 500		Int	4.4	9.5 (0.1)	-	91.0	0.0	0.1 (0.0)	-	92.5	0.0	0.0 (0.0)	-	87.9	127.7
	Int	Add	7.5	28.4 (0.5)	-	97.7	0.0	1.8 (0.1)	-	84.3	-	-	-	-	1839.1
		Int	4.6	11.0 (0.2)	-	93.7	0.0	0.2 (0.0)	-	91.4	0.7	1.6 (0.0)	-	82.5	180.5
	True	Fit	ISB	MISE	MCI_F	MCI_B	ISB	MISE	MCI_F	MCI_B	ISB	MISE	MCI_F	MCI_B	MC
Logistic	Add	Add	9.2	26.4 (0.5)	-	96.1	0.1	1.2 (0.1)	-	90.6	-	-	-	-	29.4
n = 300		Int	8.6	26.4 (0.5)	-	96.3	0.1	1.2 (0.1)	-	92.8	0.0	0.2 (0.0)	-	89.6	29.1
	Int	Add	73.8	81.4 (0.7)	-	43.9	12.1	12.8 (0.1)	-	5.1	-	-	-	-	41.9
		Int	14.1	34.7 (0.5)	-	94.1	0.4	2.2 (0.1)	-	86.5	4.1	6.5 (0.1)	-	52.5	22.7
	Add	Add	8.7	21.5 (0.3)	-	95.2	0.1	0.7 (0.0)	-	88.5	-	-	-	-	29.6
n = 500		Int	8.4	21.5 (0.3)	-	95.3	0.1	0.7 (0.0)	-	89.6	0.0	0.1 (0.0)	-	89.3	29.4
	Int	Add	72.4	77.6 (0.5)	-	36.3	12.4	12.9 (0.1)	-	3.3	-	-	-	-	42.4

$\sigma_\delta^2=4$		β_1				β_2				γ				
True	**Fit**	**ISB**	**MISE**	MCI_F	MCI_B	**ISB**	**MISE**	MCI_F	MCI_B	**ISB**	**MISE**	MCI_F	MCI_B	**APE**
	Int	13.2	27.5 (0.4)	-	93.5	0.5	1.4 (0.1)	-	80.8	3.1	5.1 (0.1)	-	58.3	23.0