REGRESSING LONGITUDINAL RESPONSE TRAJECTORIES ON A COVARIATE

Hans-Georg Müller[1] and Fang Yao[2]

[1] *Department of Statistics, UC Davis, One Shields Ave., Davis, CA 95616*
E-mail: mueller@wald.ucdavis.edu
[2] *Department of Statistics, Colorado State University, Fort Collins*

We consider the analysis of longitudinal responses which are considered random trajectories and are sampled on a sparse and irregular design, working within the framework of functional data analysis. In this functional response setting, the nonlinear influence of a one-dimensional covariate on the response curves is often of interest. We propose a simple two-dimensional smoothing approach and study its consistency. Functional principal component analysis for sparse data, using the PACE (principal analysis through conditional expectation) algorithm has been proposed recently for dimension reduction of irregularly sampled longitudinal data, and is shown to provide useful tools for the study of the variation of the observed data around the predicted trajectories. For example, the functional principal component scores of the differences between longitudinally-predicted and covariate-predicted trajectories can be used to study goodness-of-fit. We illustrate our methods with longitudinal CD4 count data and include brief reviews of functional regression models and of PACE.

1. Introduction and Review

Functional data analysis is a field that is still in rapid development. It is concerned with data that include functional components, i.e., components that can be considered to be (random) functions. Usually these functions are assumed to be observed either as continuous random trajectories, or on a dense grid with added measurement errors. Recent developments in functional data analysis include extensions of various functional regression models. A characterizing feature of functional regression models is that one of the predictors or responses is a function. The functional components are typically assumed to correspond to the realization of a stochastic process.

2   *Hans-Georg Müller and Fang Yao*

Overviews of various aspects of functional regression can be found for example in the books by Ramsay and Silverman (2002, 2005), where many other approaches in addition to those discussed below can be found, and in a review article by Rice (2004). More detailed review of some of the approaches that are only briefly mentioned in this article can be found in Müller (2005). In the following we give a brief overview of the various functional regression models that have been considered in the literature to date.

1.1. *Brief review of models with functional predictors*

The classical linear functional regression models include a functional component as a predictor and are direct extensions of the corresponding multivariate counterparts (multiple regression or multivariate regression). Frequently considered models are the functional linear regression models with predictor process X,

$$E(Y|X) = \mu + \int X(s)\beta(s)\,ds$$

for a scalar response Y, and

$$E(Y(t)|X) = \mu(t) + \int X(s)\beta(s,t)\,ds$$

for a functional response $Y(t)$. The functions β are the regression parameter functions, while μ or $\mu(t)$ is the mean of the responses. Basic ideas for such models go back to Grenander (1950) while statistical features and analysis are the focus of papers by Ramsay and Dalzell (1991), Cardot, Ferraty and Sarda (1999, 2003) and Cardot *et al.* (2003), among others. Various estimation approaches and optimal theoretical results for the case of a scalar response have been recently investigated by Hall and Horowitz (2005). Crucial differences in regard to the asymptotics between estimation of the regression function β and predicting observations Y are highlighted in Cai and Hall (2005).

   One basic problem for statistical estimation in functional regression is that the estimation of the regression parameter function β is an inverse problem, while the corresponding operators are not directly invertible. Uniqueness and existence of solutions for functional regression problems are therefore not guaranteed in general. Under mild conditions, a unique solution exists in the image space of the operator. The inversion problem in function space has been discussed in He, Müller and Wang (2000). When

estimating the regression parameter function or other characteristics from actual data, regularization is a must. General theory for regularizing estimated operators was recently developed in Hoffmann and Reiss (2005) and such developments may lead to a more comprehensive theory of regularization in functional regression.

Two main approaches have emerged with regard to implementing the regularization step: The first approach is to project on a functional basis. For the functional basis, especially the eigenbasis of the auto-covariance operator of the predictor processes X has been studied, and the projection is obtained by truncating the number of included basis functions (Rice and Silverman, 1991). This approach has also been implemented with other orthogonal bases of the function space L^2 of square differentiable functions, ranging from the traditional Fourier basis to other orthogonal function systems, such as the wavelet basis (Morris et al., 2003) or a suitably restricted family of B-splines, for example with a fixed number of knots (Shi, Weiss and Taylor, 1996; Rice and Wu, 2000). A second common approach to regularization is based on penalized likelihood or penalized least squares. To date, this approach has been implemented via penalized smoothing splines (Cardot, Crambes and Sarda 2005), ridge regression (Hall and Horowitz, 2005), and P-splines (Yao and Lee, 2005). Other forms of regularization are conceivable and will present interesting avenues of research.

Several variants have been proposed to extend the basic functional regression models with functional predictor to more complex models. These variants are motivated by the corresponding extensions of the classical multivariate regression model towards more general regression models, and include:

(1) The extension from one single index on the right hand side of the functional linear model to the case of several single indices, or projections. This is an extension analogous to projection pursuit regression (Friedman and Stuetzle, 1981; Lingjærde and Liestøl, 1998) and the multiple-index model (Chiou and Müller, 2004) and was proposed by James and Silverman (2005), who named it Functional Adaptive Model Estimation.

(2) Another variant is the generalized functional linear model (James, 2002; Cardot and Sarda, 2005; Müller and Stadtmüller, 2005). Here the response is a scalar Y which belongs to a general distribution family such as the exponential family. This extension corresponds to extending the classical linear regression model to a generalized linear model for the case of functional data. The generalized functional linear model can be described

as

$$E(Y|X) = g\left(\alpha + \int \beta(t)X(t)\,dt\right),$$

where α is an intercept parameter in the functional linear predictor and g a suitably chosen link function. The variance of the responses typically is a function of the mean, either through an exponential family assumption as in the generalized linear model, or through an otherwise postulated variance function via quasi-likelihood, in which case no exponential family distribution needs to be specified. A nonparametric quasi-likelihood version where both link function and variance function are estimated nonparametrically is included in the approach of Müller and Stadtmüller (2005).

(3) The case of sparse and irregular noisy data as they are encountered in longitudinal studies; more about this below. For the model with sparse functional predictors and sparse functional responses such an extension was discussed in Yao, Müller and Wang (2005a) and for the case of the generalized functional linear model with sparse predictors in Müller (2005).

(4) Varying-coefficient functional models, for example as the domains of the prediction functions increase, have been considered in the form of a "historical linear functional model" as proposed by Malfait and Ramsay (2003) or of a varying-coefficient regression model discussed in Müller and Zhang (2005), where the goal was the nonparametric analysis for mean residual lifetime as outcome Y. The latter paper also contains a proposal for functional quantile regression based on functional principal components analysis. An alternative approach for functional quantile regression based on penalized splines is discussed in Cardot, Crambes and Sarda (2005).

1.2. *Brief review of models with functional response*

Another class of functional regression models that is of particular interest in many medical, biological or social sciences applications concerns the case where the response is a random function and the predictor a scalar or vector. This is often the situation in longitudinal studies where individuals are followed over time and repeated measurements are made of a quantity of interest. The goal then is to study the relationship between the functional responses and some subject-specific covariate or measurement. Such models have various applications (Faraway, 1997) and have been referred to as Functional Response Models in Chiou, Müller and Wang (2004), which also includes a review of models with functional response.

We assume in the following the covariate X is scalar. Since the models of interest are generally nonparametric, in order to avoid the curse of dimension, it is necessary to include a dimension reduction step if the covariate is multivariate. Dimension reduction can be achieved through an additive model, a single index model or other methods; a model for response curves based on multiple single indices with unknown link functions and quasi-likelihood was proposed in Chiou, Müller and Wang (2003). A special case of a functional response model is the product surface model $E(Y(t)|X = x) = \mu(t)\rho(x)$ for a function ρ such that $E(\rho(X)) = 1$, so that $E(Y(t)) = E[E(Y(t)|X)] = \mu(t)$. This simple approach was found to be useful for applications to medfly data in Chiou et al. (2003), and has been referred to as functional multiplicative effects model.

Given a sample of response functions $Y(t)$, let $\mu(t) = E(Y(t))$ and define the auto-covariance operator $A(f)(t) = \int f(s)\text{cov}\{X(s), X(t)\}\, ds$. We assume existence of the quantities discussed. The operator A has orthonormal eigenfunctions ψ_k and ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$. Then the process Y can be represented through the Karhunen-Loève expansion (Karhunen, 1946)

$$Y(t) = \mu(t) + \sum_{k=1}^{\infty} A_k\, \psi_k(t), \tag{1}$$

where the expansion coefficients A_k are uncorrelated random variables (r.v.s), known as functional principal component scores. They satisfy the properties $E(A_k) = 0$ and $\text{var}(A_k) = \lambda_k$, such that $\sum_k \lambda_k < \infty$ and

$$A_k = \int (Y(t) - \mu(t))\psi_k(t)\, dt. \tag{2}$$

These developments are a direct extension of the corresponding spectral decomposition of a multivariate random vector. A crucial difference between multivariate and functional data is that only for the latter, order and neighborhood relations matter, while for the former, permutation of the components of a sample of random vectors does not affect the results of the analysis. This crucial difference makes it possible and indeed necessary to use smoothing methods when dealing with functional data.

With a covariate X, our focus is the regression

$$E(Y(t)|X = x) = \mu(t) + \sum_{k=1}^{\infty} E(A_k|X = x)\, \psi_k(t), \tag{3}$$

which is a bivariate function of t and x. For regularization of the functional data, we consider here projection onto the truncated eigenbase, where we

would include K components in the sum on the right hand side. When one deals with densely sampled data with errors or the case where entire trajectories are available, several methods have been described to obtain the functional principal components of $Y(t)$; given a sample $Y_i(t)\, i = 1, \ldots, n$, of fully observed smooth trajectories, one can simply obtain

$$\hat{\mu}(t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t), \quad \widehat{\mathrm{cov}}(Y(t), Y(s)) = \frac{1}{n} \sum_{i=1}^{n} (Y_i(t) - \hat{\mu}(t))(Y_i(s) - \hat{\mu}(s))$$

and these estimates are easily seen to converge at the $n^{-1/2}$ rate. By using perturbation methods for linear operators in Hilbert space, these results can be extended to the convergence of the resulting eigenfunction and eigenvalue estimates.

Then one usually discretizes the estimated covariance surface $\widehat{\mathrm{cov}}(Y(t), Y(s))$ to obtain the corresponding multivariate spectral decomposition. The resulting eigenvectors are smoothed or interpolated to give the eigenfunction estimates $\hat{\psi}_k$ along with the eigenvalue estimates $\hat{\lambda}_k$. The number K of components to include is often obtained in the same way as in multivariate analysis, by the scree plot or by the fraction of variance explained which is $X(t) = \mu(t) + \sum_{k=1}^{K} \tilde{\xi}_k \, \phi_k(t)$ (Mardia, Kent and Bibby, 1979). Similar results are possible if the observations are subject to additional independent errors. Regarding inference, only very few results are currently available. Besides the bootstrap which is based on resampling entire trajectories, a promising approach is Fan and Lin (1998).

1.3. *Overview*

In section 2, the focus is on the functional response model where the data are sparse and irregular, as is common in longitudinal studies. An important situation in applications is the presence of a non-functional covariate. The principal analysis through conditional expectaion (PACE) method for sparse longitudinal data is briefly reviewed. A simple smoothing procedure for regressing longitudinal responses on a covariate is introduced in section 3 and its consistency is discussed. The difference between longitudinally predicted and covariate predicted trajectories gives rise to an analogue of residuals, which in this case correspond to residual trajectories. Residual trajectories themselves can be subjected to a functional principal component analysis and can be used for goodness-of-fit checking. A small scale simulation study and an application to longitudinal CD4 measurements can be found in section 4.

2. The Functional Approach to Longitudinal Responses

In longitudinal studies, one does normally not observe complete or densely recorded trajectories and therefore the methods described above will not work. Data from such studies may be viewed as noisy measurements of a subject-specific smooth time course that are typically made at irregular and often sparse time points for a sample of individuals or items. A common approach in longitudinal data analysis is to assume parametric shapes for the trajectories and to fit random coefficient models (Diggle et al., 1994). However, this is often not entirely satisfactory when knowledge about the shape of the individual trajectories is limited, as for example in an exploratory analysis.

More flexible methods that are applicable for such situations can be based on regularized B-splines (providing essentially a flexible parametric model) as proposed by Shi, Weiss and Taylor (1996), Rice and Wu (2000) and James, Hastie and Sugar (2000) or on regularized eigenfunction expansions (Principal Analysis by Conditional Expectation (PACE); Yao, Müller and Wang, 2005b). These methods require additional restrictions, typically that the underlying processes are Gaussian. The reason for this additional restriction is that missing information must be imputed; in the principal analysis through conditional expectation (PACE) method this is done explicitly in a conditional expectation step, where it is used that the conditional expectations are linear under Gaussian assumptions.

We briefly review here the PACE method (compare Müller, 2005, for a more detailed review). The sparseness of the data is reflected by postulating a random number N_i of measurements T_{ij} for the i-th subject where the T_{ij} are i.i.d. The observations of the process Y are then Y_{ij},

$$Y_{ij} = Y(T_{ij}) + \epsilon_{ij} = \mu(T_{ij}) + \sum_{k=1}^{\infty} A_{ik}\psi_k(T_{ij}) + \epsilon_{ij}, \qquad (4)$$

where the additional measurement errors ϵ_{ij} are i.i.d. and independent of all other r.v.s, with $E\epsilon_{ij} = 0$, $\text{var}(\epsilon_{ij}) = \sigma^2$. In the PACE method, the mean function $\mu(t)$ is obtained by smoothing the scatterplot $\{(T_{ij}, Y_{ij}) : 1 \leq i \leq n, 1 \leq j \leq N_i\}$, with a local linear smoother to obtain $\hat{\mu}$; and the covariance surface via $C_{ijk} = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{ik} - \hat{\mu}(T_{ik}))$, $j \neq k$ and two-dimensional smoothing in the scatterplot $\{(T_{ij}, T_{il}; C_{ijk}) : 1 \leq i \leq n, 1 \leq j \neq k \leq N_i\}$, where the diagonal is omitted as it is contaminated by the additional error variance σ^2. This results in covariance estimates $\widehat{\text{cov}}(Y(t), Y(s))$ for all applicable t, s, whence one proceeds as described above for the dense case.

A detailed analysis of the convergence properties of the eigenvalue/eigen-function estimates $(\hat{\lambda}_k, \hat{\psi}_k)$ under sparse and less sparse situations can be found in Hall, Müller and Wang (2005). While in the dense case one obtains the rates $n^{-1/2}$ for the eigenfunction estimates, this deteriorates to the nonparametric rate $n^{-2/5}$ for twice differentiable eigenfunctions in the sparse case.

Besides estimating the components (eigenfunctions, eigenvalues and the error variance σ^2) for which consistent estimators are readily available, a major goal when applying FDA to sparse longitudinal data is the prediction of individual trajectories. To achieve such predictions, a sensible idea is to substitute estimates for the terms that appear in the Karhunen-Loève expansion,

$$\hat{Y}(t) = \hat{\mu}(t) + \sum_{k=1}^{K} \hat{A}_k \, \hat{\psi}_k(t). \tag{5}$$

Here the estimates $\hat{A}_k$ for the functional principal components A_k have not been defined yet. As has been shown in Yao *et al.* (2005b), substituting estimates based on the integral representation (2) for functional principal component scores does not work for sparse data, due to the large approximation error that one incurs in that case. Hence, alternative estimates for A_k are needed.

Such alternative estimates can be found under joint Gaussian assumptions on the random processes Y and the errors ϵ_{ij}. With $\widetilde{Y}_i = (Y_{i1}, \cdots, Y_{iN_i})^T$, $\mu_i = (\mu(T_{i1}), \cdots, \mu(T_{iN_i}))^T$, $\psi_{ik} = (\psi_k(T_{i1}), \cdots, \psi_k(T_{iN_i}))^T$ and

$$(\Sigma_{Y_i})_{j,l} = \text{cov}(Y_i(T_{ij}), Y_i(T_{il})) + \sigma^2 \delta_{jl},$$

where $\delta_{jl} = 1$ if $j = l$ and $= 0$ otherwise, one can easily calculate

$$E[A_{ik}|Y_{i1}, \ldots, Y_{iN_i}] = \lambda_k \psi_{ik}^T \Sigma_{Y_i}^{-1}(\widetilde{Y}_i - \mu_i). \tag{6}$$

Estimates for the quantities on the r.h.s. are available from the procedures described above, leading to estimates for the predicted functional principal component scores,

$$\widehat{E}[A_{ik}|Y_{i1}, \ldots, Y_{iN_i}] = \hat{\lambda}_k \hat{\psi}_{ik}^T \widehat{\Sigma}_{Y_i}^{-1}(\widetilde{Y}_i - \hat{\mu}_i). \tag{7}$$

When plugging these estimates into (5), we refer to the resulting trajectories as the estimated *longitudinally predicted trajectories*, given by

$$\hat{Y}_i(t) = \hat{\mu}(t) + \sum_{k=1}^{K} \widehat{E}[A_{ik}|Y_{i1}, \ldots, Y_{iN_i}] \, \hat{\psi}_k(t). \tag{8}$$

These estimates target the actual longitudinally predicted trajectories, given by

$$\tilde{Y}_i(t) = E(Y_i(t)|Y_{i1}, \ldots, Y_{iN_i}) = \mu(t) + \sum_{k=1}^{K} E(A_k|Y_{i1}, \ldots, Y_{iN_i})\,\psi_k(t). \quad (9)$$

Statistical properties of estimates (7), (8) of longitudinally predicted trajectories and their convergence to the actual longitudinally predicted trajectories were investigated in Yao *et al.* (2005b). Due to the conditioning step, for sparse longitudinal data, the longitudinally predicted trajectories are distinct from the actual individual trajectories $Y_i(t)$. If the N_i (number of measurements per subject) are bounded r.v.s, which reflects the actual situation in most longitudinal studies, longitudinally predicted trajectories will not converge to the true trajectories, which therefore are out of reach. However, convergence to the true individual trajectories will occur if a condition such as $\min_{1 \le i \le n} N_i \to \infty$ holds. It follows from the arguments in Müller (2005), p. 229, that under further regularity conditions one may then obtain $E(A_k|Y_{i1}, \ldots, Y_{iN_i}) \to \int Y(t)\psi_k(t)\,dt = A_k$, as $n \to \infty$, which implies that estimated longitudinally predicted trajectories are consistent for the actual individual trajectories as the number of measurements made per subject increases.

3. Predicting longitudinal trajectories from a covariate

Estimates for longitudinally predicted trajectories as discussed in the previous section are obtained by conditioning on the sparse measurements obtained for a subject whose trajectory is to be predicted. These estimates are based on the principle of borrowing strength from the entire sample of measurements by pooling the data for all subjects for estimating the mean function, the eigenfunctions and the eigenvalues. A functional data analysis approach, although nonparametric, is therefore very different from an approach where the functions are estimated separately for each individual. Such an approach would not be very efficient.

A covariate, if available, plays no role so far in predicting the longitudinal trajectories. We consider now the problem of predicting a trajectory in the presence of a covariate. We refer to these predictions as *covariate predicted trajectories*, to emphasize their distinction from the *longitudinally predicted trajectories* that were defined previously and for which a covariate is not taken into account.

Analogous to (3) and (9), the covariate predicted trajectory at covariate

level x and time t is defined as

$$\tilde{\mu}(x,t) = E(Y(t)|X = x) = \mu(t) + \sum_{k=1}^{K} E(A_k|X = x)\,\psi_k(t). \qquad (10)$$

Note that $\tilde{\mu}$ is a surface in x and t. Since in the sparse case we do not have consistent estimates of A_k available, the method proposed in Chiou *et al.* (2003) for regressing response functions on a covariate cannot be immediately extended; in particular, substituting $E(A_k|Y_{i1}, \ldots, Y_{iN_i})$ for A_k is not feasible, as in general, $E[E(A_k|Y_{i1}, \ldots, Y_{iN_i})|X] \neq E(A_k|X)$. As an alternative, we propose here to aggregate the sparse longitudinal data with similar covariate levels and then to smooth the resulting aggregated scatterplots.

The available data sample consists of n i.i.d. observations $(X_i, Y_{i1}, \ldots, Y_{iN_i})$, $i = 1, \ldots, n$. Here X_i is a one-dimensional random covariate, and the Y_{ij} are the available sparse measurements for the i-th subject, while N_i is the number of measurements. The N_i are assumed to be i.i.d. with finite moments; an example would be that they are Poisson random variables. The idea for the prediction is simple: Given a covariate level x for which the predicted trajectory is to be calculated, we select a window $[x - h, x + h]$ with bandwidth h, assemble those data for which the associated covariate X_i falls into this neighborhood, and then apply a local linear smoother to the resulting scatterplot which is indexed by time, i.e., $\{(T_{ij}, Y_{ij}) : X_i \in [x-h, x+h]\}$. Essentially, this corresponds to a smoothed surface in t and x, which is implemented by first aggregating the data whose covariate values are near x, followed by a one-dimensional smoothing step in the time direction.

Formally, to fit the weighted local least squares smoother at a fixed point t with bandwidth b (Fan and Gijbels, 1996), one chooses a nonnegative symmetric kernel function K and minimizes

$$\sum_{i:|X_i-x|\leq h} \sum_{j=1}^{N_i} K\{(t - T_{ij})/b\}\,(Y_{ij} - [a_0 + a_1(t - T_{ij})])^2$$

w.r.t. parameters a_0, a_1. The explicit solution to this weighted least squares problem is well known to be a linear estimator in the data that can be written as a weighted average with smoothing weights $w_{ij}(t)$,

$$\hat{\mu}(x,t) = \sum_{i:|X_i-x|\leq h} \sum_{j=1}^{N_i} w_{ij}(t)Y_{ij} = \hat{a}_0(t) = \frac{P_{02}(t)\,P_{10}(t) - P_{01}(t)\,P_{11}(t)}{P_{00}(t)\,P_{02}(t) - P_{01}(t)^2} \quad (11)$$

where

$$P_{0r}(t) = \sum_{i:|X_i-x|\le h} \sum_{j=1}^{N_i} (t-T_{ij})^r \, K\{(t-T_{ij})/h\},$$

$$\text{and } P_{1r}(t) = \sum_{i:|X_i-x|\le h} \sum_{j=1}^{N_i} (t-T_{ij})^r \, Y_{ij} \, K\{(t-T_{ij})/h\},$$

and the $w_{ij}(t)$ are implicitly defined weight functions.

This estimator is linear in the data Y_{ij} and depends on the choice of bandwidths h and b. In practice, one-curve-leave-out cross-validation (Rice and Silverman, 1991) is a good option: For each $1 \le i_0 \le n$, one omits the data $(X_{i_0}, Y_{i_01}, \ldots, Y_{i_0 N_{i_0}})$ in turn, and uses the remaining sample to obtain one-curve-leave-out predictions $\hat{Y}^{(-i_0)}(x,t)$. Then (h,b) are chosen so as to minimize the sum of squared prediction errors (SPE)

$$SPE = \sum_{i_0=1}^{n} \sum_{j=1}^{N_{i_0}} (Y_{i_0 j} - \hat{Y}^{(-i_0)}(X_{i_0}, T_{i_0 j}))^2.$$

To discuss consistency of this approach, we assume the regularity conditions (A1), (A3), (A4), (A6) and (B2) of Yao *et al.* (2005b) and additionally assume that the covariate level X_i is independent of number N_i and timing T_{ij} of measurements, furthermore that all joint and marginal densities of (X_i, T_{ij}, Y_{ij}) are twice differentiable and uniformly bounded away from 0 and ∞ on a domain $(x,t) \in S$ where the consistency is desired, furthermore that second moments of all r.v.s exist and are uniformly bounded. We also assume that the target function $\tilde{\mu}(\cdot,\cdot)$ (10) is twice continuously differentiable with uniformly bounded derivatives.

Theorem. *Under the above assumptions, if $h \to 0$, $b \to 0$ and $nhb^2 \to \infty$ as $n \to \infty$, it holds that*

$$\sup_{(x,t)\in S} |\hat{\mu}(x,t) - \tilde{\mu}(x,t)| = O_p\left(h^2 + b^2 + \frac{1}{(nh)^{1/2}b}\right).$$

We only give a brief sketch of the proof. Using representation (11), one first establishes that $\sum_{i:|X_i-x|\le h} \sum_{j=1}^{N_i} w_{ij}(t) = 1$ and $\sum_{i:|X_i-x|\le h} \sum_{j=1}^{N_i} w_{ij}(t)(X_i - x) = O_p(h^2)$, uniformly in t. Then we de-

12   *Hans-Georg Müller and Fang Yao*

compose

$$\sup_{(x,t)\in S} |\hat{\mu}(x,t) - \tilde{\mu}(x,t)| \leq \sup_{(x,t)\in S} |\sum_{i:|X_i-x|\leq h}\sum_{j=1}^{N_i} w_{ij}(t)\{Y_{ij} - \tilde{\mu}(x,t)\}|$$

$$\leq \sup_{(x,t)\in S} |\sum_{i:|X_i-x|\leq h}\sum_{j=1}^{N_i} w_{ij}(t)\{Y_{ij} - \tilde{\mu}(X_i,T_{ij})\}|$$

$$+ \sup_{(x,t)\in S} |\sum_{i:|X_i-x|\leq h}\sum_{j=1}^{N_i} w_{ij}(t)\{\tilde{\mu}(X_i,T_{ij}) - \tilde{\mu}(x,T_{ij})\}|$$

$$+ \sup_{(x,t)\in S} |\sum_{i:|X_i-x|\leq h}\sum_{j=1}^{N_i} w_{ij}(t)\{\tilde{\mu}(x,T_{ij}) - \tilde{\mu}(x,t)\}|$$

$$= \sup_{(x,t)\in S} |I| + \sup_{(x,t)\in S} |II| + \sup_{(x,t)\in S} |III|.$$

The analysis is conditional on the T_{ij}. One easily finds $E(I) = 0$, while the terms in the sum which share the same index i are dependent. This situation is analogous to Lemma 1 in Yao *et al.* (2005b), and adapting the proof leads to $I = O_p((nh)^{-1/2}b^{-1})$. Taylor expansion of $\tilde{\mu}(X_i,T_{ij})$ around $\tilde{\mu}(x,T_{ij})$ yields $II = O_p(h^2)$, and again by a Taylor expansion one finds that $III = O(b^2)$, with bounds that are uniform on S.

Note that the proposed approach for regressing response functions on a covariate is quite simple, amounting to no more than a two-dimensional smoothing step. The smoothing step can be implemented with any available two-dimensional smoother. In the following illustrations, we choose an implementation based on weighted local least squares.

4. Illustrations

The estimated longitudinally predicted trajectories $\hat{Y}_i(t)$ are closest to the actual trajectories since they incorporate actual measurements from the trajectories, albeit sparse ones. These trajectories can be used to check the goodness-of-fit of the covariate-predicted trajectories $\hat{\mu}(x,t)$. For this purpose we define the differences

$$R_i(t) = \hat{Y}_i(t) - \hat{\mu}(x,t), \tag{12}$$

which assume a role similar to the residuals in ordinary regression analysis. The analysis of these residual trajectories, extending the usual residual analysis, can be done by subjecting the residual trajectories to a functional principal component analysis.

We first describe a small simulation study which confirms that indeed the longitudinally-predicted trajectories are generally closer to actual trajectories than the covariate-predicted trajectories, and then proceed to illustrate the methods with longitudinal CD4 data.

4.1. *A simulation comparison of longitudinally-predicted and covariate-predicted trajectories*

The simulation aims at comparing the difference between actual individual trajectories and predicted trajectories that one obtains either with estimated longitudinally predicted trajectories (8) or with estimated covariate-predicted trajectories $\hat{\mu}(X_i, t)$ (11), given the covariate level X_i for the i-th subject.
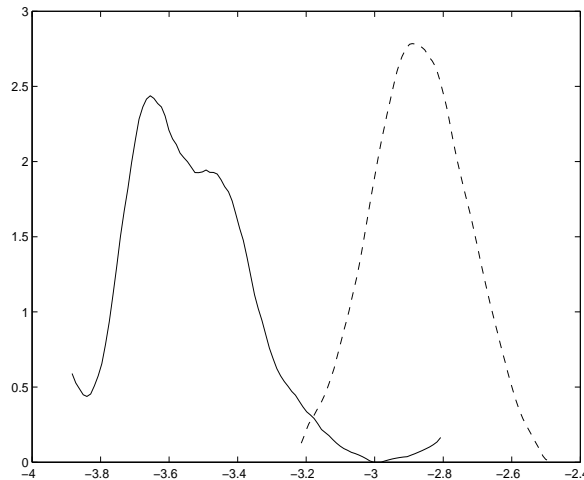

Fig. 1.   Densities of the distribution of the logarithm of mean integrated relative errors (13), obtained for 200 simulated samples, comparing the prediction errors for longitudinally-predicted (9) and covariate-predicted (11) trajectories. Density estimate for errors of longitudinally-predicted trajectories is the solid curve, while for the covariate-predicted trajectories it is the dashed curve.

The underlying simulation model for the ith subject is $Y_i(t) = \mu(t) + \sum_{k=1}^{2} A_{ik}\phi_k(t)$, where $\mu(t) = t + \sin(t)$, $\phi_1(t) = -\cos(\pi t/10)/\sqrt{5}$, and $\phi_2(t) = \sin(\pi t/10)/\sqrt{5}$, $0 \leq t \leq 10$. The random coefficients A_{ik} are generated as follows: The covariate $X_i \stackrel{\text{i.i.d.}}{\sim} N(0,1)$, $A_{i1} = g_1(X_i) + B_{i1}$

14   *Hans-Georg Müller and Fang Yao*

and $A_{i2} = g_2(X_i) + B_{i2}$, where $B_{ik} \overset{\text{i.i.d.}}{\sim} N(0,3)$. The nonlinear functions $g_1(z) = c_{11}[\sin(z) - c_{12}]$ and $g_2(z) = c_{21}(z^3 - c_{22})$ are used, where the constants $c_{11}, c_{12}, c_{21}, c_{22}$ are adjusted so that $E[g_k(X_i)] = 0$ and $\text{var}(g_k(X_i)) = 1$, $k = 1, 2$. The jth measurement made for the ith subject is given by $Y_{ij} = Y_i(T_{ij}) + \epsilon_{ij}$, where the additional measurement errors ϵ_{ij} are also assumed to be normal with mean 0 and constant variance $\sigma^2 = 1$. Each curve is sampled at a random number of points, chosen from a discrete uniform distribution on $\{2, \ldots, 6\}$, and the locations of the measurements were uniformly distributed on $[0, 10]$.

We use the procedures as described in Yao *et al.* (2005b) to estimate the model components, such as $\mu(t)$, λ_k and $\phi_k(t)$. The number of included components K in the estimates (8) for longitudinally predicted trajectories was chosen by AIC, based on a pseudo-likelihood, as described in that paper. In the implementation of the estimates for covariate-predicted trajectories $\hat{\mu}(X_i, t) = \widehat{E}\{Y(t)|X = X_i\}$ (11), we used subjective bandwidth choices.

Table 1. Lower, median, and upper quartiles of mean integrated relative error (13), obtained from 200 samples, comparing longitudinally-predicted (9) and covariate-predicted (11) trajectories as estimates for actual trajectories.

Method	Lower	Median	Upper
Longitudinally-predicted	.026	.028	.032
Covariate-predicted	.052	.056	.062

The criterion used to measure the discrepancy between the actual observed individual trajectories $Y_i(t)$ and the estimated trajectories from either method is the mean integrated relative error (MIRE), defined as follows,

$$\text{MIRE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\int\{Y_i(t) - \widehat{Y}_i(t)\}^2 dt}{\int Y_i^2(t) dt},$$

$$\text{or} \qquad \text{MIRE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\int\{Y_i(t) - \hat{\mu}(X_i, t)\}^2 dt}{\int Y_i^2(t) dt}. \qquad (13)$$

The lower, median and upper quartiles of the criterion MIRE (13) obtained from 200 simulation samples using the two estimation methods are reported in Table 1. Density estimates for the values of the MIRE (13) in

the logarithmic scale for the two approaches are shown in Figure 1. These simulation comparisons clearly show that longitudinally-predicted trajectories are considerably more accurate than covariate-predicted trajectories, justifying the definition of residuals R_i as in (12).

4.2. *Application to longitudinal CD4 data*

These longitudinal data have been collected for 283 homosexual men who became HIV positive between 1984 and 1991. They consist of irregularly measured CD4 percentages with 1 to 14 measurements per subject after seroconversion, with a median count of 7 measurements. These data were analyzed previously with functional data analysis methodology in Yao *et al.* (2005b) and have also been analyzed by other authors. Details about the design of the study can be found in Kaslow *et al.* (1987).
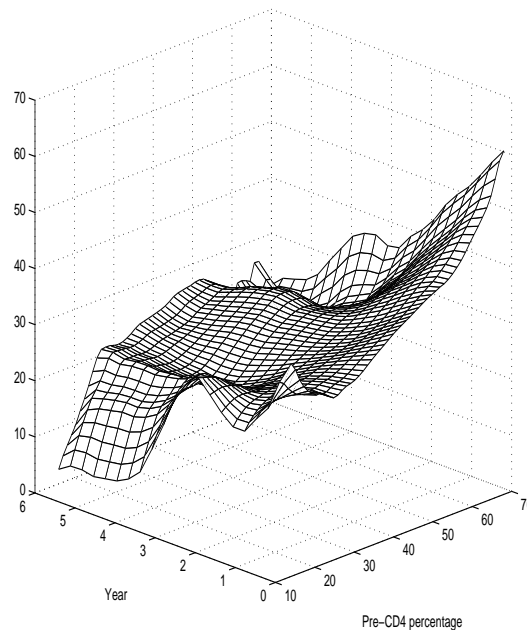

Fig. 2.   Covariate-predicted trajectories (10), estimated via (11), as a function of time t in the study (in years) and Pre-CD4 level x.

In a first step, we applied the methods described in Yao et al. (2005b),

16   *Hans-Georg Müller and Fang Yao*

estimating the model components of the longitudinal CD4 process, such
as mean function, covariance surface and the eigenvalues/eigenfunctions.
In a second step, we then obtained the trajectory estimates $\widehat{Y}_i(t)$ (8) by
principal analysis through conditional expectation (PACE); these estimates
correspond then to the estimated longitudinally predicted trajectories.

As covariate X_i we used Pre-CD4 percentage, which is a time-
independent measurement that is available for each patient at the time
of entry into the longitudinal study. We then investigated the influence of
Pre-CD4 percentage on the shape of the longitudinal trajectories. As de-
scribed above, two-dimensional smoothing is applied to obtain the estimates
of covariate-predicted trajectories $\hat{\mu}(x,t)$ (11) that target the true covariate-
predicted trajectories $\tilde{\mu}(x,t)$ (10). Viewed for all (x,t) simultaneously, the
$\hat{\mu}(x,t)$ form a surface that is shown in Figure 2. With few exceptions, the
predicted trajectories are generally decreasing in t. The rate of decrease
appears to be related to the Pre-CD4 level, higher levels being associated
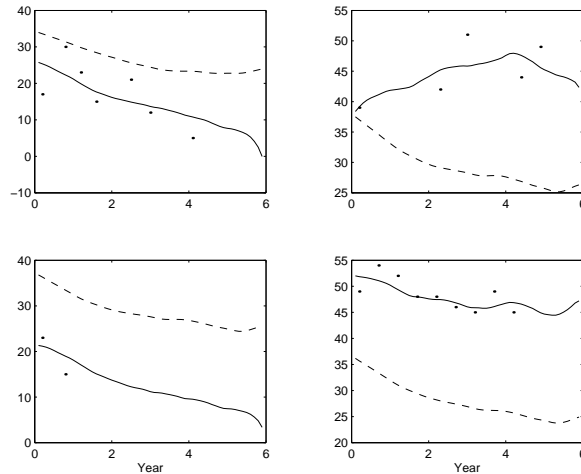with higher rates of decrease.

Fig. 3.   Observed measurements of CD4 percentage (dots) and estimated longitudi-
nally predicted CD4 profiles (solid curves) as well as covariate-predicted profiles (dashed
curves) for four randomly selected subjects.

Estimated CD4 profiles comparing the estimates obtained for both lon-
gitudinally and covariate predicted trajectories (8) and (11) for four ran-
domly selected patients are shown in Figure 3. As expected from the simu-

lation results, the longitudinally predicted trajectories (solid lines) are seen to be considerably closer to the observed measurements. The residuals (12) are displayed in the left panel of Figure 4. These residual profiles can be subjected to a functional principal component analysis in order to study the behavior of these functional residuals.
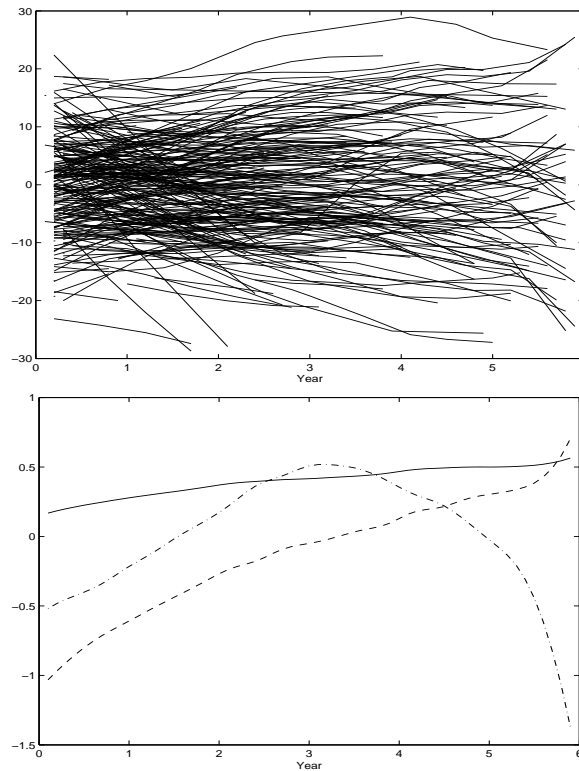

Fig. 4.   Left: Residual profiles obtained by subtracting covariate-predicted trajectories from longitudinally-predicted trajectories for the CD4 data. Right: First three eigenfunction of the residual profiles (first eigenfunction, solid; second eigenfunction, dashed; third eigenfunction, dash-dotted).

The first three eigenfunction resulting from such an eigen-analysis of residuals (the AIC criterion led to the selection of three components) are shown in the right panel of Figure 4. The variation of the residual profiles near the endpoints is of interest. If one wants to study lack of fit (bias) for any proposed fitting method, the usual goodness-of-fit plot known from

18   *Hans-Georg Müller and Fang Yao*

multiple linear regression, plotting residuals versus the predicted values,
is not easily feasible here since both residuals and predicted values are
trajectories. Using functional principal component scores for the residual
trajectories, one can for example plot the first two scores of the residu-
als against the covariate value (CD4 percentage) itself. This plot is shown
in Figure 5 and does not provide evidence for systematic lack of fit for
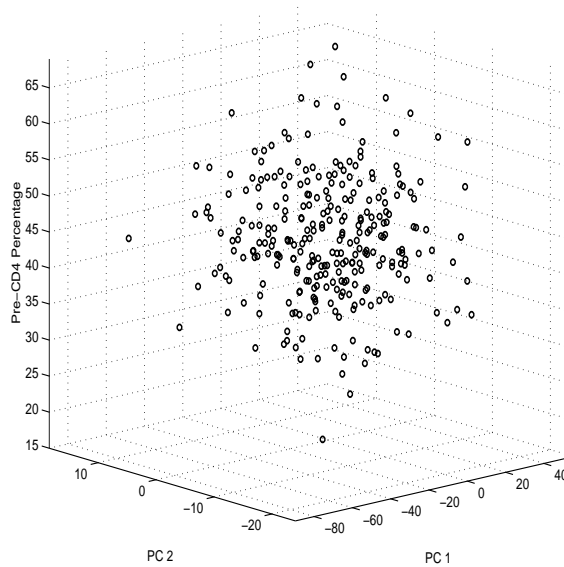covariate-predicted trajectories.

Fig. 5.   Relationship between first (x-axis) and second (y-axis) functional principal com-
ponent score of the residual profiles and the Pre-CD4 percentage (z-axis).

Acknowledgments

References

Cai, T., Hall, P. (2005). Prediction in functional linear regression. Preprint.
Cardot, H., Crambes, C., Sarda, P. (2005). Quantile regression when the covari-
ates are functions. Preprint.

Cardot, H., Ferraty, F., Mas, A., Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Statist.* **30**, 241-255.

Cardot H., Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *J. Multiv. Analysis* **92**, 24-41.

Chiou, J.M., Müller, H.G. (2004). Quasi-likelihood regression with multiple indices and smooth link and variance functions. *Scand. J. Statist.* **31**, 367-386.

Chiou, J. M., Müller, H. G., Wang, J. L. (2003). Functional quasi-likelihood regression models with smooth random effects. *J. Royal Statist. Assoc. Series B* **65**, 405-423.

Chiou, J.M., Müller, H.G., Wang, J.L. (2004). Functional response models. *Statistica Sinica* **14**, 675-693.

Chiou, J.M., Müller, H.G., Wang, J.L., Carey, J.R. (2003). A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statistica Sinica* **13**, 1119-1133.

Fan, J., Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* Chapman , Hall, London.

Fan, J., Lin S. K. (1998). Test of significance when data are curves. *J. Amer. Statist. Assoc.* **93**, 1007-1021.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics* **39**, 254-262.

Friedman, J.H., Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76**, 817-823.

Grenander, U. (1950). Stochastic processes and statistical inference. *Arkiv för Matematik*, 195-276.

Hall, P., Horowitz, J.L. (2005). Methodology and convergence rates for functional linear regression. Preprint.

Hall, P., Müller, H.G., Wang, J.L. (2005). Properties of principal component methods for functional and longitudinal data analysis. Preprint.

He, G., Müller, H.G., Wang, J.L. (2000). Extending correlation and regression from multivariate to functional data. *Asymptotics in Statistics and Probability*, Ed. Puri, M.L., VSP International Science Publishers, pp. 301-315.

Hoffmann, M., Reiss, M. (2005). Nonlinear estimation for linear inverse problems with errors in the operator. Preprint.

James, G. (2002). Generalized linear models with functional predictors. *J. Royal Statist. Soc. B* **64**, 411-432.

James, G., Hastie, T. G., Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587-602.

James, G., Silverman, B.W. (2005). Functional adaptive model estimation. *J. Amer. Statist. Assoc.* **100**, 565-576

Karhunen, K. (1946). Zur Spektraltheorie stochastischer Prozesse. *Ann. Acad. Sci. Fennicae* **A I 37**.

Lingjærde, O. C., Liestøl, K. (1998). Generalized projection pursuit regression. *SIAM J. Sci. Comput.* **20**, 844-857.

Malfait N., Ramsay J. O. (2003). The historical functional linear model. *Canadian Journal of Statistics* **31**, 115-128.

Mardia, K. V., Kent, J. T., Bibby, J. M. (1979). *Multivariate Analysis*, London:

20   *Hans-Georg Müller and Fang Yao*

Academic Press.

Morris, J.S, Vannucci, M., Brown, P.J., Carroll, R.J. (2003). Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis (with discussion). *J. Amer. Statist. Assoc.* **98**, 573-597.

Müller, H.G. (2005). Functional Modelling and Classification of Longitudinal Data. *Scand. J. Statist.* **32**, 223-240.

Müller, H.G., Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.* **33**, 774-805.

Müller, H.G., Zhang, Y. (2005). Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics* **59**, 676-685.

Ramsay, J., Dalzell, C.J. (1991). Some tools for functional data analysis. *J. Royal Statist. Soc. Series B* **53**, 539-572.

Ramsay, J., Silverman, B. (2002). *Applied functional data analysis*, New York: Springer.

Ramsay, J., Silverman, B. (2005). *Functional data analysis*, New York: Springer.

Rice, J. (2004). Functional and longitudinal data analysis: Perspectives on smoothing. *Statistica Sinica* **14**, 631-647.

Rice, J., Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Royal Statist. Soc. Series B*, **53**, 233-243.

Rice, J., Wu, C. (2000). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253-259.

Shi, M., Weiss, R. E., Taylor, J. M. G. (1996). An analysis of paediatric CD4 counts for Acquired Immune Deficiency Syndrome using flexible random curves. *Applied Statistics*, **45**, 151-163.

Yao, F., Lee, T. C. M. (2005). Penalized spline models for functional principal component analysis. *J. Royal Statistical Society B*, to appear.

Yao, F., Müller, H.G., Wang, J.L. (2005a). Functional linear regression analysis for longitudinal data. *Ann. Statist.*, to appear.

Yao, F., Müller, H.G., Wang, J.L. (2005b). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577-590.