Functional mixture regression

FANG YAO *

Department of Statistics, University of Toronto, Toronto, Ontario M5S 3G3, Canada
fyao@utstat.toronto.edu

YUEJIAO FU

Department of Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada

THOMAS C. M. LEE

Department of Statistics, University of California, Davis, CA 95616, USA

Summary

In functional linear models (FLMs), the relationship between the scalar response and the functional predictor process is often assumed to be identical for all subjects. Motivated by both practical and methodological considerations, we relax this assumption and propose a new class of functional regression models that allow the regression structure to vary for different groups of subjects. By projecting the predictor process onto its eigenspace, the new functional regression model is simplified to a framework that is similar to classical mixture regression models. This leads to the proposed approach named as functional mixture regression (FMR). The estimation of FMR can be readily carried out using existing software implemented for functional principal component analysis and mixture regression. The practical necessity and performance of FMR are illustrated through applications to a longevity analysis of female medflies and a human growth study. Theoretical investigations concerning the consistent estimation and prediction properties of FMR along with simulation experiments illustrating its empirical properties are presented in the supplementary material available at Biostatistics online. Corresponding results demonstrate that the proposed approach could potentially achieve substantial gains over traditional FLMs.

Keywords: Dimensional reduction; Eigenfunction; Functional data; Functional linear model; Functional principal components; Mixture regression; Smoothing.

1. Introduction

Recently, there has been an increased interest in regression models for functional data. In the simplest setting, the functional predictor and the scalar response are related by a linear operator. Given a scalar response $Y$ on $\hat{R}$ and a smooth random predictor process $X(\cdot)$ on a compact support $\hat{T}$ that is square integrable (i.e., $\int_{\hat{T}} X^2(t)dt < \infty$), the classical functional linear model (FLM) relates $Y$ and $X$ by

$$E(Y|X) = \int_{\hat{T}} \beta(t)X(t)dt,$$

(1.1)
where the regression parameter function $\beta(\cdot)$ is also assumed to be smooth and square integrable. See Ramsay and Silverman (2005) for a comprehensive introduction. For further theoretical studies on (1.1), see Cardot and others (1999, 2003), Cai and Hall (2006) and Hall and Horowitz (2007).

Driven by the needs of generalizing the basic linear relationship in (1.1), several extensions of the above FLM have been proposed. This is similar to, for example, extending the classical linear regression models to generalized linear models. One of the early examples is generalized FLMs (Müller and Stadtmüller, 2005), and other extensions include varying-coefficient functional models (Fan and Zhang, 2000; Fan and others, 2003) and wavelet-based functional models (Morris and others, 2003).

In line with these extensions and motivated by the fact that, due to some unknown reasons or unobserved covariates, the subjects may belong to different mutually exclusive groups that possess different mechanisms to produce the response, we propose a new class of functional regression models. Our approach achieves this goal by allowing individuals from different groups to have distinct regression functions. To be specific, denote the unknown number of groups as $K$, and let $\beta_k(t)$ be the regression function for the $k$th group, $k = 1, \ldots, K$. Then, we propose the following model:

$$E(Y|X) = \int_T \beta_k(t)X(t)dt \text{ if the subject belongs to the } k\text{th group. (1.2)}$$

We shall first illustrate the utility of our proposal through an analysis of the biodemographic characteristics of female medflies (Müller and Zhang, 2005). This study concerns the dependence of longevity on the dynamics of the early fertility process and we shall show that our proposal sheds new light on various important scientific issues. The second example, derived from the Berkeley growth study, considers the regression of heights at maturity age on the childhood growth patterns. We shall illustrate how distinct regression relations emerge and reveal the underlying gender groups, even when we were completely blinded from gender information throughout the analysis. Our proposed method can also be potentially useful in various medical applications, such as when $X(t)$ is a longitudinal biomarker and $Y$ is a disease indicating variable; for example, glomerular filtration rate in kidney diseases and postload glucose in type 2 diabetes.

Extending the classical FLM (1.1)–(1.2) is parallel to extending the classical linear regression to mixture regression (DeSarbo and Cron, 1988), thus termed as functional mixture regression (FMR). We emphasize that a main goal of FMR is to specify an appropriate functional model that is capable of identifying potentially different regression structures. This general idea can readily be adopted to various applications in which functional regression techniques are needed. We also remark that FMR is conceptually different from existing approaches for curve-based clustering (Gaffney and Smyth, 2003; James and Sugar, 2003; Luan and Li, 2003, among others). These latter methods focus on clustering the trajectories themselves, while FMR focuses on detecting the possible existence of different regression relations.

The rest of this paper is organized as follows. In Section 2, we provide a complete description of FMR and demonstrate that its estimation can be achieved using existing software implemented for functional principal component analysis (FPCA) and mixture regression. Applications of the proposed method to the above-mentioned real examples are presented in Section 3. Concluding remarks are offered in Section 4, while simulations illustrating the empirical performance and theoretical investigation on consistent estimation and prediction are deferred to supplementary material available at Biostatistics online for conciseness.

2. Functional mixture regression

We begin with the classical FLM (1.1) and review a key methodology for dimension reduction and regularization of functional data, namely, FPCA. For introductory material on FPCA, see Rice and Silverman (1991), James and others (2001), Ramsay and Silverman (2005), Yao and others (2005), among others.
2.1 FLM and eigenbasis representation

We begin with the classical FLM (1.1), where the regression function $\beta(t)$ is the same for all subjects under consideration. It is known that the functional linear operator in (1.1) is compact and not directly invertible, and thus regularization is needed and can be achieved through a truncated basis representation. In this article, we shall adopt the eigenbasis representation for reasons to be given below.

The process $X$ with finite covariance possesses a sequence of orthonormal eigenfunctions $\{\phi_m\}_{m=1,2,\ldots}$, which form a complete basis of the functional space, with associated nonnegative and nondecreasing eigenvalues $\{\lambda_m\}_{m=1,2,\ldots}$. By the well-known Karhunen–Loève expansion, the predictor process $X$ admits

$$
X(t) = \mu(t) + \sum_{m=1}^{\infty} \xi_m \phi_m(t), \quad \text{where} \quad \xi_m = \int_X \{X(t) - \mu(t)\} \phi_m(t) dt.
$$

(2.1)

The random variables $\xi_m$s are the functional principal component (FPC) scores of $X$, which are uncorrelated and satisfy $E(\xi_m) = 0$ and $\text{var}(\xi_m) = \lambda_m$, $\sum_m \lambda_m < \infty$. The trajectory $X_i$ is an i.i.d. realization of $X$ for the $i$th subject related to the response $Y_i$, $i = 1, \ldots, n$. Here, we highlight the following advantages of (2.1) for regression regularization that will also be carried over to FMR. First, the eigenfunctions are orthonormal in the $L^2$ space and the FPC scores are uncorrelated random variables, which provide both analytical and computational convenience. Second, the eigenbasis are determined by the data and will efficiently capture the dominant modes of variation. The eigenvalues often decrease rapidly and thus the infinite-dimensional predictor process can be well approximated by a small number of FPCs. This suggests a simple way to achieve regularization by truncating eigenbasis based on the total variation explained up to a threshold.

Recall that the regression parameter function $\beta$ is square integrable and $\{\phi_m\}_{m=1,2,\ldots}$ form a complete orthonormal basis, we have $\beta(t) = \sum_{m=1}^{\infty} b_m \phi_m(t)$, and hence model (1.1) can be expressed equivalently as

$$
E(Y|X) = \int_T \beta(t) \mu(t) dt + \int_T \left\{ \sum_{m=1}^{\infty} b_m \phi_m(t) \right\} \left\{ \sum_{m=1}^{\infty} \xi_m \phi_m(t) \right\} dt
$$

$$
= b_0 + \sum_{m=1}^{\infty} b_m \xi_m,
$$

(2.2)

where the intercept is $b_0$ and the coefficients are given by $b_m = \int_T \beta(t) \phi_m(t) dt$. One can see that the orthonormality of the complete eigenbasis plays a critical role in transforming the functional regression structure into a linear combination of the uncorrelated FPC scores that serve as predictor variables in (2.2).

2.2 Model specification of functional mixture regression

This section provides a complete mathematical formulation of FMR. Recall that FMR allows the predictor trajectories $X_i$ to partition into $K$ mutually exclusive groups, with each group having its own regression function $\beta_k(t)$ for producing the response $Y_i$. This idea was previously expressed by model (1.2) and it is useful to rewrite it in a similar manner as (2.2). In this paper, $K$ is unknown and will be chosen by some statistical model selection criterion addressed later. Write $\tilde{S} = \{1, \ldots, n\}$ and define the index set

$$
\tilde{C}_k = \{i \in \tilde{S}: \text{the } i\text{th subject belongs to the } k\text{th group}\}, \quad k = 1, \ldots, K.
$$

Thus, $\bigcup_{k=1}^K \tilde{C}_k = \tilde{S}$ and $\tilde{C}(k_1) \cap \tilde{C}(k_2) = \emptyset$ for $1 \leq k_1 \neq k_2 \leq K$. Write $b_{k0} = \int_T \beta_k(t) \mu(t) dt$ and $b_{km} = \int_T \beta_k(t) \phi_m(t) dt$. By analogy to (2.2), we have

$$
\beta_k(t) = \sum_{m=1}^{\infty} b_{km} \phi_m(t), \quad t \in T, \quad k = 1, \ldots, K,
$$

(2.3)
and the FMR model (1.2) can be written as

\[ E(Y_i | X_i, i \in \tilde{C}_k) = b_{k0} + \sum_{m=1}^{\infty} b_{km} \tilde{\zeta}_{im}. \] (2.3)

We note that in (2.3) the FPC scores \( \tilde{\zeta}_{im} \) serve as the predictor variables and that the infinite-dimensional feature is inherent to the functional data. The estimation of model (2.3) thus requires regularization for the predictor process \( X \), and we achieve this by truncating the infinite sum to a finite sum of \( M \) terms. Since often the eigenvalues of \( X \) decrease to zero rapidly, it is reasonable to assume that an appropriate \( M \) will be chosen that leads to a flexible yet parsimonious model. A simple and practical strategy is to examine the total variation explained up to certain threshold \( \tau \), \( M = \min \{ t : \sum_{m=1}^{t} \lambda_m / \sum_{m=1}^{\infty} \lambda_m \geq \tau \} \).

With this truncation, model (2.3) is refined,

\[ E(Y_i | X_i, M, i \in \tilde{C}_k) = b_{k0} + \sum_{m=1}^{M} b_{km} \tilde{\zeta}_{im}, \] (2.4)

as the underlying model and the regression parameter functions that we aim for are

\[ \beta_{k,M}(t) = \sum_{m=1}^{M} b_{km} \phi_m(t) \quad \text{for all } t \in \mathcal{T} \text{ and } k = 1, \ldots, K. \] (2.5)

From (2.4), the FMR model is now reduced to a form similar to classical mixture of linear regression models.

To complete the mathematical description of the FMR model (2.4) from which the statistical inference is based on, let \( \pi_k \) be the probability that a randomly selected subject is from the \( k \)th group and define \( \sigma^2_{kY} = \text{var}(Y_i | X_i) \) if \( i \in \tilde{C}_k \). Write \( \mathbf{y} = (y_1, \ldots, y_n)^T, \mathbf{b}_0 = (b_{10}, \ldots, b_{K0})^T, \mathbf{b}_k = (b_{k1}, \ldots, b_{kM})^T, \varpi = (\pi_1, \ldots, \pi_{K-1})^T, \) and \( \sigma^2_Y = (\sigma^2_{1Y}, \ldots, \sigma^2_{KY})^T. \) Denote \( \varPsi = (b_0^T, b_1^T, \ldots, b_K^T, \pi^T, \sigma_Y^T) \), the parameter space \( \Theta \) is then given by

\[ \Theta = \left\{ \varPsi : \pi_k > 0, \sum_{k=1}^{K} \pi_k = 1, \sigma^2_{kY} > 0 \text{ for all } k = 1, \ldots, K \right\}, \] (2.6)

which is an open subset of \( \tilde{R}^{(M+3)K-1}. \) Further, write \( \mathbf{\xi}_i = (\zeta_{i1}, \ldots, \zeta_{iM})^T \) for \( i = 1, \ldots, n; \) these vectors \( \mathbf{\xi}_i \)'s of the FPC scores are i.i.d. realizations of a random vector \( \mathbf{\xi} \) whose density is \( f(\mathbf{\xi} | \Lambda) \) with \( E(\mathbf{\xi}) = \mathbf{0} \) and \( \text{cov}(\mathbf{\xi}, \mathbf{\xi}) = \text{diag}(\lambda_1, \ldots, \lambda_M) = \Lambda \). The conditional density of \( Y \) given \( \mathbf{\xi} \) is

\[ f(y | \mathbf{\xi}, \varPsi) = \sum_{k=1}^{K} \pi_k f(y | \mathbf{\xi}, \mathbf{b}_k, \lambda_k, \sigma_{kY}^2), \] (2.7)

where \( f(y | \mathbf{\xi}, \mathbf{b}_k, \lambda_k, \sigma_{kY}^2) \) is the conditional density for the \( k \)th component. In general, the component density can be derived from a location-scale family (Hennig, 2000) or an exponential family (Wedel and DeSarbo, 1995) that generates identifiable mixtures. This includes most commonly adopted distributions, such as normal, gamma, exponential, Poisson, binomial, and multinomial. To emphasize the main idea of coupling functional data, we focus on the mixture of linear regressions with normal errors. It is conceptually straightforward to extend our proposal to the mixture of generalized linear models with estimation procedure modified accordingly. Note that the above formulation holds for model (2.4) and the dependence on \( M \) is suppressed for.

Since the \( \mathbf{\xi}_i \)'s are random, (2.4) becomes a random design regression model. As to be described in Section 2.3, the matrix \( \Lambda \) will be estimated in a way that is functionally independent of \( \varPsi \), therefore, the inference of (2.4) is completely based on the conditional density \( f(y | \mathbf{\xi}, \varPsi) \). As \( \lambda_1 \geq \cdots \geq \lambda_M > 0 \) for
any $M$, it is easy to see that $f(\xi|\Lambda)$ does not have all its mass in up to $K$ of $(M - 1)$-dimensional linear subspaces. This implies that the FMR model (2.4) is identifiable in the following sense (Hennig, 2000): for any 2 parameters $\psi$ and $\psi^*$ with a given predictor variable $\xi_i$, if

$$
\sum_{k=1}^{K^*} \pi_k f(y|\xi_i, b_{k0}, b_k, \sigma_{k2}^2) = \sum_{k=1}^{K^*} \pi_k^* f(y|\xi_i, b_{k0}^*, b_k^*, \sigma_{k2}^*)
$$

(2.8)

is true for all $i$ and all possible values of $y$, then $K = K^*$ and $\psi = \psi^*$ up to a permutation. It is noteworthy that an attractive property of (2.4) is that its predictor variables $\xi_{im}$s are uncorrelated, hence devoid of collinearity.

It is important to note that the condition that $X_i$ are i.i.d. processes does not exclude the possibility of $\xi_i$ itself following a mixture distribution because the Karhunen–Loève expansion (2.1) that FPC analysis is based on only requires the existence of covariance function. The inference on the regression structure is based on the conditional density (2.7) and does not depend on the distribution of $\xi_i$. For instance, suppose that $\xi_i$s follow a mixture density. The regression structure does not necessarily vary across the groups of subjects partitioned by the distribution of $\xi_i$. In other words, the assignments of group membership for an individual in FMR is determined merely by the relationship between $X_i$ and $Y_i$. Another noteworthy remark is that, for the reason of detecting different regression structures, the mixing proportion $\pi_k$ is assumed independent of the predictor $X_i$, which is similar to classical mixture regression models. This is distinct from the class of hierarchical mixture of experts arising from the neural network literature (Jiang and Tanner, 1999), where it is common to assume that the $\pi_k$s depend on covariates to allow for flexible approximation of the overall mean response function.

### 2.3 Model estimation and implementation

This subsection discusses approaches for estimating the unknowns in (2.4) and (2.5), which can be naturally done in 2 stages. Briefly, in the first stage, we perform FPCA to obtain estimates for $\phi_m$ and $\xi_{im}$, while in the second stage, these estimates are plugged into (2.4) and (2.5) for the estimation of the remaining parameters.

In practice, the observed data are noisy measurements $U_{ij}$ taken at $t_{ij}$,

$$
U_{ij} = X_i(t_{ij}) + \varepsilon_{ij} = \mu(t_{ij}) + \sum_{m=1}^{\infty} \xi_{im} \phi_m(t_{ij}) + \varepsilon_{ij}, \quad t_{ij} \in \mathcal{T},
$$

(2.9)

for $i = 1, \ldots, n$ and $j = 1, \ldots, n_j$. The measurement errors $\varepsilon_{ij}$ are assumed independent of $\xi_{im}$ with mean zero and a constant variance $E\varepsilon_{ij}^2 = \sigma_\varepsilon^2$, while a nonconstant variance function could also be assumed to account for heteroscedasticity (Yao and Lee, 2006). We first apply the principal analysis by conditional estimation (PACE) procedure of Yao and others (2005) to these noisy measurements to carry out FPCA. When this is done, the following estimates of model components are obtained: $\hat{\mu}$, $\hat{G}$, $\hat{\phi}_m$, $\hat{\lambda}_m$, $\hat{\xi}_{im}$, $m = 1, \ldots, M$. Here, $M$ is the number of FPCs that can be chosen by pseudo-Akaike Information Criterion or other related selectors or simply as the minimum number of FPCs that explain a sufficiently large proportion of the total variation for the predictor process. We adopted the latter approach and found that the 90% threshold works excellently for our numerical examples. In general, one may need to navigate several choices of the threshold values to determine the model that provides an adequate fit with parsimonious structure.

For conciseness, we refer to Yao and others (2005) for a complete description of the FPCA technique used in this paper. Here, we only present the integral and PACE estimates of the FPC scores $\xi_{im}$ with the
notation introduced in Section 2.1. The integral estimate is given by

\[ \hat{\xi}_m = \sum_{j=2}^{n_i} (U_{ij} - \hat{\mu}(t_{ij})) \hat{\phi}_m(t_{ij})(t_{ij} - t_{i,j-1}), \]  

(2.10)

which is motivated by the definition of the FPC scores as inner products; that is, \( \xi_{im} = \int_T [X_i(t) - \mu(t)] \phi_m(t) \, dt \). For the PACE estimates, write \( \mu_i = (\mu(t_{i1}), \ldots, \mu(t_{in_i}))^T \), \( U_i = (U_{i1}, \ldots, U_{in_i})^T \), and \( \phi_{lm} = (\phi_m(t_{i1}), \ldots, \phi_m(t_{in_i}))^T \), and let the \((j, l)\)-th entry of the \( n_i \times n_i \) matrix \( \Sigma_{kl} \) be \( (\Sigma_{kl})_{j,l} = G(t_{ij}, t_{il}) + \sigma_j^2 \delta_{jl} \) with \( \delta_{jl} = 1 \) if \( j = l \) and 0 if \( j \neq l \). Substituting estimates for \( \mu_i, \lambda_m, \phi_{lm} \), and \( \Sigma_{kl} \), we have the PACE estimates

\[ \hat{\xi}_m = \hat{\lambda}_m \hat{\phi}_m^T \Sigma_{U_i}^{-1} (U_i) - \hat{\mu}_i. \]  

(2.11)

It is widely known that when the design points \( t_{ij} \) are dense, the traditional integral estimates of the FPC scores \( \xi_{im} \), denoted by \( \hat{\xi}_{im} \), and is given by (2.10) below, are usually satisfactory. By contrast, the PACE estimates \( \hat{\xi}_m \) as in (2.11) is more suitable when the design points are moderate or sparse. Corresponding software is available at http://www.utstat.toronto.edu/fyao.

Once the crucial estimates \( \hat{\xi}_im \) of the FPC scores are obtained by either (2.10) or (2.11), the regression coefficients \( b_{km} \) in (2.4) can be estimated in a relatively straightforward manner: with \( \hat{\xi}_im \) as the predictor variables, then \( b_{km} \) can be estimated by standard mixture regression estimation method (e.g., expectation maximization-based method). The fitted FMR model and regression functions are then given by, for \( k = 1, \ldots, K \),

\[ \hat{E}(Y_i|X_i, M) = \hat{b}_{k0} + \sum_{m=1}^M \hat{b}_{km} \hat{\xi}_m, \quad \hat{b}_{kM}(t) = \sum_{m=1}^M \hat{b}_{km} \phi_m(t) \quad \text{if } i \in \mathcal{C}_k. \]  

(2.12)

Notice that the estimated FPC scores \( \hat{\xi}_im \) are unique up to a sign change related to the direction of the estimated eigenfunctions \( \hat{\phi}_m(t) \). This property is also carried over to the estimated regression coefficients \( \hat{b}_{km} \). For the choice of \( K \), one could apply any well-studied model selection criterion, and we adopt the Bayesian information criterion (BIC) that has provided good results in a variety of applications of model-based clustering (e.g., see Fraley and Raftery, 2002, and references therein).

For conducting inference procedures on the regression function(s), we could exploit nonparametric bootstrap methods with a suitable label-switching strategy for mixture regression to avoid nonidentifiability of component labels. More specifically, we first resample all the individuals with replacement to obtain a bootstrap sample, \((U_{i1}^b, \ldots, U_{in_i}^b, y_i^b); i = 1, \ldots, n, b = 1, \ldots, B, \) and perform the FPCA step, where \( B \) is the number of bootstrap replicates. Then the estimated FPC scores \((\hat{\xi}_{i1}^b, \ldots, \hat{\xi}_{in_i}^b); i = 1, \ldots, n \) are fed into a mixture regression model with \( K \) components, where \( M^b \) is chosen in FPCA using the same criterion as for the original sample, and \( K \) is the number of mixture components selected for the original sample. To correctly label the latent groups, we examine the distances from the estimated regression functions of the bootstrap sample, \( \hat{\beta}_e^b(t) = \sum_{m=1}^{M^b} \hat{b}_{km}^b \phi_e^b(t) \), to those obtained from the original sample, that is, \( \hat{\beta}_e^b = \arg\min_{\beta_e^b} \int_T (\hat{\beta}_e^b(t) - \hat{\beta}_e(t))^2 \, dt \) for each \( k = 1, \ldots, K \). We note that this bootstrap procedure also provides evidence for model identifiability.

We have derived theoretical results in terms of consistency of model estimation and prediction for FMR. In establishing such results, a first technical difficulty encountered is the fact that the estimates of the regression functions in the FMR model (2.4) are based on the estimated FPC scores \( \hat{\xi}_im \) not on the “true” \( \xi_{im} \). Thus, existing theories of mixture regression models are no longer applicable. Another major challenge is due to the lack of analytic expressions for \( \hat{b}_{km} \). Therefore, customary theoretical arguments previously used in FLMs cannot be applied. Due to space limitation, these technical contents such as the relevant theorems, assumptions, auxiliary lemmas, and proofs to the supplementary material available at Biostatistics online.
3. Applications

3.1 Longevity and early fertility of Mediterranean flies

To illustrate the need of the proposed approach, we analyze the egg-laying data from a fertility study conducted for 1000 female medflies as described in Carey and others (1998). Our goal is to determine the dependence of longevity of the medflies on their early fertility process. One of the basic questions of evolutionary theory is to what extent lifespan is driven by enabling increased reproduction. Diverting resources used for maintenance and repair into reproductive activity may shorten lifespan (Partridge and Harvey, 1985; Westendorp and Kirkwood, 1999). The selected sample of 139 medflies includes those that were fertile during an early life period defined by the first 20 days and also survived beyond. The trajectories corresponding to the number of daily eggs during this early life period constitute the functional predictors, while remaining lifetime serves as the response that is an important proxy for longevity and quantifying the evolutionary fitness of individual flies. As a preprocessing step to achieve homogeneity, a log-transform of egg counts was applied.

These predictor trajectories (obtained by applying the PACE algorithm in FPCA step) are shown in the left panel of Figure 1. Most egg-laying trajectories display a rise toward a time of peak fertility followed by a decline. There is substantial variation in the steepness of the rise to the various maximal level of egg-laying and also in the timing of the peak and the rate of decline. The smooth estimate of the mean fertility function is also displayed, while the estimates of the first 2 eigenfunctions are shown in the right panel, explaining 76.8% and 14.5% of the total variation of the trajectories, respectively. These eigenfunctions reflect the modes of variation (Castro and others, 1986) and the dynamics of predictor processes. Two components were chosen, and they account for more than 90% of the variation in the data, that is, $\tau = 0.9$.

It is of interest to identify shape changes in early life reproductive trajectories that tend to influence evolutionary longevity. To conduct an adequate analysis, we would inspect whether the regression relationship varies due to some unknown mechanism. It is noticed that there is no obvious grouping effect in the predictor trajectories observed. This can be seen from the perspective of the estimated FPC scores (right panel of Figure 1) that are often viewed as subject-specific summaries. However, when the remaining lifetimes are graphed versus the FPC scores in the right panel of Figure 2, the lifespan seems driven by the early fertility differently with considerably longer lifetimes for some flies whose predictor patterns (in the left panel) might be similar. To verify this conjecture, we applied the FMR approach and unsurprisingly 2 mutually exclusive groups with different regression structures were suggested by BIC ($K = 2$),
Fig. 2. Left: estimated FPC scores for the 139 flies with different markers (circles and crosses) used for representing the 2 mutually exclusive groups with different regressions as detected by FMR. Right: responses $Y_i$ (remaining lifetime in day) against estimated FPC scores for the 139 flies with corresponding group assignments.

Fig. 3. Top panels: estimated regression functions $\hat{\beta}_1$ (solid in top left) and $\hat{\beta}_2$ (solid in top right) of the 2 groups detected by FMR along with 95% bootstrap confidence bands (dashed). Bottom panels: predictor trajectories of the flies (indicated by circles in the right panel of Figure 2) that correspond to $\hat{\beta}_1$ (bottom left) and of the flies (indicated by crosses) that correspond to $\hat{\beta}_2$ (bottom right).

where different markers (circles and crosses) were used for enhanced visualization of such phenomenon in the right panel of Figure 2.

The estimates of regression functions $\beta_1$ (solid) and $\beta_2$ (solid) serving as weighting functions shown in the top panels of Figure 3 indicate how the lifespan is influenced by the early fertility process, depending on which group a fly belongs to. We applied the nonparametric bootstrap procedure as described in Section 2.3 and constructed the 95% bootstrap confidence bands by taking 2.5th and 97.5th quantiles of
1000 replicates shown in the same panels. This provides a measure of accuracy of our point estimation and evidence for model identifiability as well. For illustration, we plotted again the predictor trajectories of 2 groups separately in bottom panels of Figure 3, where the flies in the middle panel \((n_1 = 32)\) correspond to \(\beta_1\) that have longer lifetimes (indicated by circles in the right panel of Figure 2) and those in the right panel \((n_2 = 107)\) to \(\beta_2\) those live shorter (indicated by crosses). Overall higher level of fertility seems to shorten lifespan. More specifically, if a fly belongs to the group according to \(\beta_1\), a slow rise to a lower peak of egg production helps to prolong the lifespan. By contrast, the flies in the other group with larger reproductivity around day 15 often have shorter lifetimes. These findings shed some new insight by distinguishing distinct underlying mechanisms relating longevity and early fertility. This may help experimenters look into evolutionary interpretation and implication of these mechanisms for different medflies.

We conclude this example by a comparison with a FLM, where the leave-one-subject-out cross-validated relative prediction errors \(\text{CVRPE} = \sum_{i=1}^{n}(Y_i - \hat{Y}_{(-i)}^2)/\sum_{i=1}^{n}Y_i^2\) were obtained for FMR as 0.163 and for FLM as 0.372, indicating a substantial gain of 56% in prediction ability.

### 3.2 Berkeley growth study

Studies of human growth dynamics are an important topic in biological and medical applications that have profound impact for many years. This example concerns the Berkeley growth data originally published in Tuddenham and Snyder (1954) and analyzed by Ramsay and others (1995) in terms of height acceleration to reveal the dynamics of human growth. Similar data, for example, the Zurich growth data, were also studied from this perspective using various smoothing approaches (Gasser and others, 1984, among others). It is known that the growth patterns of boys and girls during their pubertal spurts differ significantly in terms of magnitude and timing. Mainly for demonstration purpose, in this example, we study the human growth from a different perspective by examining the dependence of the height at maturity age 18 (scalar response) on the dynamic pattern till age 9 (predictor process) before pubertal spurts (see Figure 4).

![Fig. 4. Left and middle: height trajectories from age 1 to 9 for 39 boys (left) and 54 girls (middle). Right: smooth estimates of the first (solid) and second (dashed) eigenfunctions, accounting for 88.7% and 9.8% of total variation.](http://biostatistics.oxfordjournals.org/)

Downloaded from http://biostatistics.oxfordjournals.org/ by guest on November 7, 2016
The data analyzed consist of height records for 39 boys and 54 girls, where the measurements were taken quarterly from ages 1 to 2, annually from 2 to 8, and semiannually from 8 till 18. It is worth mentioning that, for illustration, we shall blind ourselves from the gender information throughout the analysis, that is, the gender is an unknown or hidden factor that we expect the proposed approach is capable to detect.

We first carried out FPCA for the predictor process by pooling 93 trajectories together. Shown in the right panel of Figure 4 are the smooth estimates of the first 2 eigenfunctions account for 88.7% and 9.8% of the total variation, respectively, where the first eigenfunction is in the direction of the overall trend and the second shows a contrast between early and late times. The estimated FPC scores displayed in the top left panel of Figure 5 do not show a strong separation between boys and girls. However, when we examined the plot of the response $Y_i$ against the estimated FPC scores in the top right panel, the separation between boys and girls becomes more apparent (circles for boys and crosses for girls). This phenomenon seems to suggest different regression relations for each gender.

The FMR approach indeed worked beautifully and led to 2 distinct groups based on BIC, that is, $K = 2$. Moreover, the partition based on FMR results corresponds to the gender group as expected in which only 1 boy and 2 girls were misclassified when we inspected the cross-validated classification. The regression functions for boys and girls are displayed in the bottom panels of Figure 5 along with the 95% bootstrap confidence bands. Recall that the first eigenfunction is an overall shift. The common

![Fig. 5. Top left: the estimated FPC scores obtained by the integral method (2.10) for boys (circles) and girls (crosses). Top right: responses $Y_i$ (heights at maturity) against estimated FPC scores for boys (circles) and girls (crosses). Bottom panels: estimated regression functions (solid) of 2 groups that correspond to boys (bottom left) and girls (bottom right) along with 95% bootstrap confidence bands (dashed).]
increasing trends indicate more weights on the height measurements at later times. It is expected and also
confirmed by the data that boys are usually taller than girls at maturity due to the increasing patterns of
individual predictor trajectories and the faster ascending regression weights after around age 5. Again by
comparison with a FLM based on the CVRPEs, 0.0005 for FMR and 0.0017 for FLM, we observed a
substantial reduction of 70%.

4. CONCLUDING REMARKS

In this paper, we investigated a new type of functional regression models, FMR, that relate a scalar re-
sponse to an infinite-dimensional predictor process through possibly different regression structures. The
proposed FMR is particularly useful when the use of a single regression structure for modeling all sub-
jects is inadequate. The need for this modeling approach was demonstrated through 2 real data examples
as well as simulation studies that can be found in the online Appendix. Utilizing FPCA as a means for
regularization caused by the infinite-dimensional nature of the predictor process, we developed a simple
and yet flexible framework that is similar to classical mixture regression with a set of uncorrelated FPC
scores as predictors. The estimation procedures can be easily implemented with existing softwares for
FPCA and mixture regression. Lastly, we note that the proposed modeling framework can be immediately
extended to nonnormal mixture settings and/or to other nonlinear link functions.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

The authors are grateful to the reviewer and the associate editor for many helpful and constructive com-
ments, which led to a significantly improved version of the paper. Conflict of Interest: None declared.

FUNDING

Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to F.Y., Y.F.;
National Science Foundation (0707037 and 1007520) to T.C.M.L.

REFERENCES

11–22.
Sinica 13, 571–591.
of fecundity to mortality, longevity, and lifetime reproduction in a large, cohort of mediterranean fruit fly females.
Journal of Gerontology: Biological Sciences and Medical Sciences 53, B245–251.
Journal of Classification 5, 249–282.


[Received June 17, 2010; revised September 20, 2010; accepted for publication September 21, 2010]