Functional Linear Regression Via Canonical Analysis[1]

Guozhong He[2], Hans-Georg Müller[3], and Jane-Ling Wang[3]

November 2007

ABSTRACT

We study regression models for the situation where both dependent and independent variables are square integrable stochastic processes. Questions concerning definition and existence of the corresponding functional linear regression models and some basic properties are explored. We derive a representation of the regression parameter function in terms of the canonical components of the processes involved. This representation establishes a connection between functional regression and functional canonical analysis and leads to new approaches for functional linear regression analysis. Several specific procedures for the estimation of the regression parameter function using canonical expansions are explored. As an application example, we present an analysis of mortality data for cohorts of medflies, obtained in experimental studies of aging and longevity.

Key words: Canonical components, covariance operator, functional data analysis, functional linear model, stochastic process, parameter function, longitudinal data.

Mathematical Subject Classification: Primary 62G05, 62H25; Secondary 60G12, 62M10.

---

[1]Research supported by NSF Grants DMS03-54448, DMS04-06430 and DMS05-05537.

[2]University of California, San Francisco, Diabetes Unit.

[3]Department of Statistics, University of California, Davis

Address for correspondence: Hans-Georg Müller, Department of Statistics, University of California, One Shields Ave., Davis, CA 95616, USA; e-mail: `mueller@wald.ucdavis.edu`

1. INTRODUCTION

With the advancement of modern technology, data sets which contain repeated measurements obtained on a dense grid are becoming ubiquitous. Such functional data can be viewed as a sample of curves or functions, and are referred to as functional data. We consider here the extension of the linear regression model to the case of functional data. In this extension, both the response variable and the predictor variable are random functions rather than random vectors. It is well known (Ramsay and Dalzell, 1991, Ramsay and Silverman, 2005) that the traditional linear regression model for multivariate data, defined as

$$\boldsymbol{Y} = \boldsymbol{\alpha}_0 + \mathbf{X}^T \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}, \tag{1}$$

may be extended to the functional setting as follows. For $s \in T_1, t \in T_2$, let

$$Y(t) = \alpha_0(t) + \int_{T_1} X(s)\beta_0(s,t)ds + \varepsilon(t). \tag{2}$$

In the classical linear model (1), $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ are random vectors in $\mathbf{R}^{m_2}$, $\mathbf{X}$ is a random vector in $\mathbf{R}^{m_1}$, and $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are $m_2 \times 1$ and $m_1 \times m_2$ matrices containing the regression parameters. The vector $\boldsymbol{\varepsilon}$ has the usual interpretation of an error vector, with $E[\boldsymbol{\varepsilon}] = 0$, $\text{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I$, I denoting the identity matrix. In the more general model (2), random vectors $\mathbf{X}$ and $\mathbf{Y}$ in (1.1) are replaced by random functions defined on the intervals T_1 and T_2. The extension of the classical linear model (1) to the functional linear model (1.2) is obtained by replacing the matrix operation on the r.h.s. of (1.1) with an integral operator in (2). In the original approach of Ramsay and Dalzell (1991), a penalized least squares approach using L-splines was adopted and applied to a study in temperature-precipitation patterns based on data from Canadian weather-stations.

Model (2) for the case of scalar responses has found much recent interest (Cardot and Sarda, 2005; Müller and Stadtmüller 2005; Hall and Horowitz 2006), while the case of functional responses has been less thoroughly investigated (Ramsay and Dalzell 1991; Yao, Müller and Wang, 2005). Further discussions on various estimation procedures can be found in the insightful monograph of Ramsay and Silverman (2005). In this paper, we propose an alternative approach to predict $Y(\cdot)$ from $X(\cdot)$, by means of a novel representation of the regression parameter functions$\beta_0(s,t)$. Several interesting and distinctive features of functional linear models emerge in the development of this canonical expansion approach.

It is well known that in the classical multivariate linear model, the regression slope parameter matrix is uniquely determined by $\boldsymbol{\beta}_0 = \text{cov}(\mathbf{X})^{-1}\text{cov}(\mathbf{X}, \mathbf{Y})$, as long as the covariance matrix $\text{cov}(\mathbf{X})$ is invertible. In contrast, the corresponding parameter function $\beta_0(\cdot,\cdot)$, appearing in (2), is typically not identifiable. This identifiability issue is discussed in Section 2. It relates to the compactness of the covariance operator of the process X which renders this operator non-invertible. We show in Section 2, how restriction to a subspace allows to circumvent this problem. Under suitable restrictions, the components of model (2) are shown to be well-defined.

We then develop a novel approach to estimate the parameter function $\beta_0(\cdot,\cdot)$, by utilizing a canonical decomposition which is presented in Theorem 3.3. This decomposition links Y and X through their functional canonical correlation structure. The corresponding canonical components provide a bridge between canonical analysis and linear regression modeling. Canonical components provide a decomposition of the structure of the dependency between Y and X, and lead to a natural expansion of the regression parameter function $\beta_0(\cdot,\cdot)$, thus aiding in its interpretation. The canonical regression decomposition also leads to a new family of estimation procedures for functional regression analysis. We refer to this methodology as functional canonical regression analysis. Classical canonical correlation analysis (CCA) was introduced by Hotelling (1931) and connected to function spaces by Hannan (1961). Substantial extensions and connections to reproducing kernel Hilbert spaces were recently developed in Eubank and Hsing (2007). Canonical correlation is known not to work particularly well for very high dimensional multivariate data as it involves an inverse problem. Leurgans, Moyeed and Silverman (1993) tackled the difficult problem of extending CCA to the case of infinite-dimensional functional data and discussed the precarious regularization issues one faces, while He, Müller and Wang (2003, 2004) further explored various aspects and proposed practically feasible regularization procedures for functional CCA.

While CCA for functional data is worthwhile but difficult to implement and interpret, the canonical approach to functional regression is found here to compare well with the usual smoothed least-squares approach in an application example (Section 5). This demonstrates a potentially important role for canonical decompositions and CCA for both functional and multivariate regression analysis. The functional linear model (2) includes the varying coefficient linear model studied in Hoover, Rice, Wu, and Yang (1998) and Fan and Zhang (1998) as a special case, where $\beta(s,t) = \beta(t)\delta_t(s)$; here $\delta_t(\cdot)$ is a delta function centered at t and $\beta(t)$ is the varying coefficient function. Other forms of functional regression models with vector valued predictors and functional responses were considered by Faraway (1997), Shi, Taylor and Weiss (1997), Rice and Wu (2000) and Chiou, Müller and Wang (2003).

The paper is organized as follows. Functional canonical analysis and functional linear models for L_2-processes are introduced in Section 2. Sufficient conditions for the existence of functional normal equations are given in Proposition 2.2. The canonical regression decomposition and its properties are the theme of Section 3. In Section 4, we consider two estimation techniques to obtain parameter function estimates. This is the basic model component of interest in functional linear models, in analogy to the parameter vector in classical linear models. One proposed estimation method is based on the canonical regression decomposition, while a second procedure utilizes a more direct two-dimensional local polynomial smoothing approach. As selection criterion for tuning parameters such as bandwidths or number of canonical components, we use minimization of prediction error via one-curve-leave-out cross-validation (Rice and Silverman, 1991). The proposed estimation procedures are then applied to mortality data obtained for cohorts of medflies (Section 5). Our goal is the prediction of a random trajectory of mortality for a female cohort of flies from the trajectory of mortality for

a male cohort which was raised in the same cage. We find that the proposed functional canonical regression method is competitive in terms of one-leave-out prediction error. Additional results on canonical regression decompositions and properties of functional regression operators are compiled in Section 6. All proofs are collected in Section 7.

2. FUNCTIONAL REGRESSION AND THE FUNCTIONAL NORMAL EQUATION

In this section we explore the formal setting as well as identifiability issues for functional linear regression models. Both response and predictor functions are considered to come from a sample of pairs of random curves. A basic assumption is that all random curves or functions are square integrable stochastic processes. Consider a measure μ on a real index set T, and let $L_2(T)$ be the class of real valued functions such that $\|f\|^2 = \int_T |f|^2 d\mu < \infty$. This is a Hilbert space with the inner product $\langle f, g\rangle = \int_T fg\ d\mu$, and we write $f = g$ if $\int_T (f-g)^2 d\mu = 0$. The index set T can be a set of time points, such as $T = \{1, 2, \ldots, k\}$, a compact interval $T = [a, b]$, or even a rectangle formed by two intervals S_1 and S_2, $T = S_1 \times S_2$. We focus on index sets T that are either compact real intervals or compact rectangles in $\mathbf{R}^2$, and consider μ to be the Lebesgue measure on $\mathbf{R}^1$ or $\mathbf{R}^2$. Extensions to other index sets T and other measures are self-evident. An L_2-process is a stochastic process $X = \{X(t), t \in T\}$, $X \in L_2(T)$, with $E[\|X\|^2] < \infty$.

Let $X \in L_2(T_1)$ and $Y \in L_2(T_2)$.

Definition 2.1 *A functional linear model is defined as:*

$$Y(t) = \alpha_0(t) + \int_{T_1} X(s)\beta_0(s,t)ds + \varepsilon(t), \tag{3}$$

where $\beta_0 \in L_2(T_1 \times T_2)$ is the parameter function, $\varepsilon \in L_2(T_2)$ is the random error process with $E[\varepsilon(t)] = 0$ for $t \in T_1$, and ε and X are uncorrelated, in the sense that $E[X(t)\varepsilon(s)] = 0$ for $s, t \in T_1$.

We assume from now on without loss of generality that all processes considered have zero mean functions, $EX(t) = 0$ and $EY(s) = 0$ for all t, s. Define the regression integral operator $\mathcal{L}_X : L_2(T_1 \times T_2) \to L_2(T_2)$, such that

$$(\mathcal{L}_X\beta)(t) = \int_{T_1} X(s)\beta(s,t)ds, \quad \text{defined for any } \beta \in L_2(T_1 \times T_2).$$

Then (3) can be rewritten as

$$Y = \mathcal{L}_X\beta_0 + \varepsilon. \tag{4}$$

Denoting the covariance functions of X and Y as:

$$r_{XX}(s,t) = \text{cov}[X(s), X(t)],\ s, t \in T_1,$$

$$r_{YY}(s,t) = \text{cov}[Y(s), Y(t)],\ s, t \in T_2,$$

$$r_{XY}(s,t) = \mathrm{cov}[X(s), Y(t)],\ s \in T_1, t \in T_2,$$

the auto-covariance operator of X is the integral operator $R_{XX} : L_2(T_1) \to L_2(T_1)$ such that

$$(R_{XX}u)(s) = \int_{T_1} r_{XX}(s,t)u(t)dt, \quad u \in L_2(T_1).$$

Replacing r_{XX} by r_{YY}, r_{XY}, we analogously define operators $R_{YY} : L_2(T_2) \to L_2(T_2)$ and $R_{XY} : L_2(T_2) \to L_2(T_1)$, similarly R_{YX}. Then, R_{XX} and R_{YY} are compact, self-adjoint and nonnegative definite operators, and R_{XY} and R_{YX} are compact operators (Conway, 1985). We refer to He et al. (2003) for a discussion of various properties of these operators.

Another linear operator of interest is the integral operator $\Gamma_{XX} : L_2(T_1 \times T_2) \to L_2(T_1 \times T_2)$,

$$(\Gamma_{XX}\beta)(s,t) = \int_{T_1} r_{XX}(s,w)\beta(w,t)\, dw. \tag{5}$$

The operator equation

$$r_{XY} = \Gamma_{XX}\beta,\ \beta \in L_2(T_1 \times T_2) \tag{6}$$

is a direct extension of the least squares normal equation and is referred to as the functional population normal equation.

Proposition 2.2 *The following statements are equivalent for a function $\beta_0 \in L_2(T_1 \times T_2)$:*

(a) β_0 *satisfies the linear model* (4);

(b) β_0 *is a solution of the functional normal equation* (6);

(c) β_0 *minimizes* $E\|Y - \mathcal{L}_X\beta\|^2$, *among all* $\beta \in L_2(T_1 \times T_2)$.

The proof is in Section 7. In the infinite dimensional case, the operator Γ_{XX} is a Hilbert-Schmidt operator in the Hilbert space L_2, according to Proposition 6.6 below. A problem we face is that it is known from functional analysis that a bounded inverse does not exist for such operators. A consequence is that the parameter function β_0 in (3), (4) is not identifiable without additional constraints. In a situation where the inverse of the covariance matrix does not exist in the multivariate case, a unique solution of the normal equation always exists within the column space of $\mathrm{cov}(\mathbf{X})$, and this solution then minimizes $E\|Y - \mathcal{L}_X\beta\|^2$ on that space. Our idea to get around the non-invertibility issue in the functional infinite dimensional case is to extend this approach for the non-invertible multivariate case to the functional case. Indeed, as is demonstrated in Theorem 2.3 below, under the additional Condition (C1) below, the solution of (6) exists in the subspace defined by the range of Γ_{XX}. This unique solution indeed minimizes $E\|Y - \mathcal{L}_X\beta\|^2$.

Condition (C1) refers to the Karhunen -Loève decompositions (Ash and Gardner, 1975) for L_2-processes X and Y,

$$X(s) = \sum_{j=1}^{\infty} \xi_j\theta_j(s),\ s \in T_1 \quad \text{and} \quad Y(t) = \sum_{j=1}^{\infty} \zeta_j\varphi_j(t),\ t \in T_2, \tag{7}$$

with random variables ξ_j, ζ_j, $j \geq 1$, and orthonormal families of L_2-functions $\{\theta_j\}_{j\geq 1}$ and $\{\varphi_j\}_{j\geq 1}$. Here $E\xi_j = E\zeta_j = 0$, $E\xi_i\xi_j = \lambda_{Xi}\delta_{ij}$, $E\zeta_i\zeta_j = \lambda_{Yi}\delta_{ij}$, and $\{(\lambda_{Xi}, \theta_i)\}, \{(\lambda_{Yj}, \varphi_j)\}$ are the eigenvalues and eigenfunctions of the covariance operators R_{XX} and R_{YY}, respectively, with $\sum_i \lambda_{Xi} < \infty$, $\sum_i \lambda_{Yi} < \infty$ Note that δ_{ij} is the Kronecker symbol with $\delta_{ij} = 1$, for $i = j$, $\delta_{ij} = 0$ for $i \neq j$.

We consider a subset of L_2 on which inverses of the operator Γ_{XX} can be defined. We note that as a Hilbert-Schmidt operator, Γ_{XX} is compact and therefore is not invertible on L_2. According to Conway (1985, p.50), the range of Γ_{XX}

$$G_{XX} = \{\Gamma_{XX}h : h \in L_2(T_1 \times T_2)\},$$

is characterized by

$$G_{XX} = \{g \in L_2(T_1 \times T_2) : \sum_{i,j=1}^{\infty} \lambda_{Xi}^{-2} |\langle g, \theta_i\varphi_j\rangle|^2 < \infty,\ g \perp \ker(\Gamma_{XX})\}, \tag{8}$$

where $\ker(\Gamma_{XX}) = \{h : \Gamma_{XX}h = 0\}$. Defining

$$G_{XX}^{-1} = \{h \in L_2(T_1 \times T_2) : h = \sum_{i,j=1}^{\infty} \lambda_{Xi}^{-1} |\langle g, \theta_i\varphi_j\rangle \theta_i\varphi_j,\ g \in G_{XX}\},$$

we find that Γ_{XX} is a one-to-one mapping from the vector space $G_{XX}^{-1} \subset L_2(T_1 \times T_2)$ onto the vector space G_{XX}. Thus restricting Γ_{XX} to a subdomain defined by the subspace G_{XX}^{-1}, we can define its inverse for $g \in G_{XX}$ as

$$\Gamma_{XX}^{-1} g = \sum_{i,j=1}^{\infty} \lambda_{Xi}^{-1} |\langle g, \theta_i\varphi_j\rangle \theta_i\varphi_j. \tag{9}$$

Then Γ_{XX}^{-1} satisfies the usual properties of an inverse in the sense that $\Gamma_{XX}\Gamma_{XX}^{-1}g = g$, for all $g \in G_{XX}$, and $\Gamma_{XX}^{-1}\Gamma_{XX}h = h$, for all $h \in G_{XX}^{-1}$.

Condition(C1) *The L_2-processes X and Y with Karhunen -Loève decompositions (7) satisfy*

$$\sum_{i,j=1}^{\infty} \left\{ \frac{E[\xi_i\zeta_j]}{\lambda_{Xi}} \right\} < \infty.$$

That the solution to the non-invertibility problem as outlined above is viable in the functional case, if (C1) is satisfied, is demonstrated by the following basic theorem of functional linear models (Theorem 2.3).

Theorem 2.3 *(Basic theorem of functional linear models) A unique solution of the linear model (4) exists in* $\ker(\Gamma_{XX})^{\perp}$ *if and only if X and Y satisfy Condition (C1). In this case, the unique solution is of the form*

$$\beta_0^*(t, s) = (\Gamma_{XX}^{-1} r_{XX})(t, s). \tag{10}$$

As a consequence of Proposition 2.2, solutions of the functional linear model (4), of the functional population normal equation (6), and minimizers of $E\|Y - \mathcal{L}_X\beta\|^2$ are all equivalent, and allow the usual projection interpretation.

Proposition 2.4 *Assume X and Y satisfy Condition (C1). Then the following are equivalent:*

(a) The set of all solutions of the functional linear model (4);

(b) The set of all solutions of the population normal equation (6);

(c) The set of all minimizers of $E\|Y - \mathcal{L}_X\beta\|^2$, for $\beta \in L_2(T_1 \times T_2)$;

(d) The set $\beta_0^* + \ker(\Gamma_{\mathrm{XX}}) = \{\beta_0^* + \mathrm{h} | \mathrm{h} \in \mathrm{L}_2(\mathrm{T}_1 \times \mathrm{T}_2), \Gamma_{\mathrm{XX}}\mathrm{h} = 0\}$.

It is well known that in a finite dimensional situation, the linear model (6) always has a unique solution in the column space of Γ_{XX}, obtained by using a generalized inverse of the matrix Γ_{XX}. However, in the infinite-dimensional case, such a solution does not always exist. The following example demonstrates that a pair of L_2 -processes does not necessarily satisfy Condition (C1). In this case, the linear model (6) does not have a solution.

Example 2.5 Assume processes X and Y have Karhunen -Loève expansions (7), where the random variables ξ_i, ζ_j satisfy

$$\lambda_{Xi} = E[\xi_i^2] = \frac{1}{i^2}, \ \lambda_{Yj} = E[\zeta_j^2] = \frac{1}{j^2}, \tag{11}$$

and let

$$E[\xi_i\zeta_j] = \frac{1}{(i+1)^2(j+1)^2}, \text{ for } i, j \geq 1. \tag{12}$$

As shown in He et al. (2003), (11) and (12) can be satisfied by a pair of L_2-processes with appropriate operators R_{XX}, R_{YY}, and R_{XY}. Then

$$\sum_{i,j=1}^{\infty}\left\{\frac{E[\xi_i\zeta_j]}{\lambda_{Xi}}\right\}^2 = \lim_{n\to\infty}\sum_{i,j=1}^{n}\left[\frac{i}{(i+1)(j+1)}\right]^4 = \lim_{n\to\infty}\sum_{i=1}^{n}\left[\frac{i}{(i+1)}\right]^4\sum_{j=1}^{\infty}\frac{1}{(j+1)^4} = \infty,$$

therefore Condition (C1) is not satisfied.

3. CANONICAL REGRESSION ANALYSIS

Canonical analysis is a time-honored tool for studying the dependency between the components of a pair of random vectors or stochastic processes; for multivariate stationary time series, its utility has been established in the work of Brillinger (1983). In this section we demonstrate that functional canonical decomposition provides a useful tool to represent functional linear models. The definition of functional canonical correlation for L_2-processes is as follows.

Definition 3.1. *The first canonical correlation ρ_1 and weight functions u_1 and v_1 for L_2-processes X and Y are defined as*

$$\rho_1 = \sup_{u\in L_2(T_1),\, v\in L_2(T_2)} \mathrm{cov}(\langle u, X\rangle, \langle v, Y\rangle) = \mathrm{cov}(\langle u_1, X\rangle, \langle v_1, Y\rangle), \tag{13}$$

where u and v are subject to

$$\mathrm{var}(\langle u_j, X\rangle) = 1, \qquad \mathrm{var}(\langle v_j, Y\rangle) = 1 \tag{14}$$

for $j = 1$. The kth canonical correlation ρ_k and weight functions u_k, v_k for processes X and Y for $k > 1$ are defined as

$$\rho_k = \sup_{u \in L_2(T_1),\, v \in L_2(T_2)} \mathrm{cov}(\langle u, X\rangle, \langle v, Y\rangle) = \mathrm{cov}(\langle u_k, X\rangle, \langle v_k, Y\rangle),$$

where u and v are subject to (14) for $j = k$, and

$$\mathrm{cov}(\langle u_k, X\rangle, \langle u_j, Y\rangle) = 0, \quad \mathrm{cov}(\langle v_k, X\rangle, \langle v_j, Y\rangle) = 0,$$

for $j = 1, \cdots, k-1$. We refer to $U_k = \langle u_k, X\rangle$ and $V_k = \langle v_k, Y\rangle$ as the kth canonical variates, and to $(\rho_k, u_k, v_k, U_k, V_k)$ as the kth canonical components.

It has been shown in He et al. (2003) that canonical correlations do not exist for all L_2-processes, but that Condition (C2) below is sufficient for the existence of canonical correlations and weight functions.

Condition (C2) Let X and Y be L_2-processes, with Karhunen -Loève decompositions (7) satisfying

$$\sum_{i,j=1}^{\infty} \left\{ \frac{E[\xi_i \zeta_j]}{\lambda_{Xi} \lambda_{Yj}^{1/2}} \right\}^2 < \infty.$$

We remark that Condition (C2) implies Condition (C1).

The proposed functional canonical regression analysis exploits features of functional principal components and of functional canonical analysis. In functional principal component analysis one studies the structure of an L_2-process via its decomposition into the eigenfunctions of its covariance operator, the Karhunen -Loève decomposition (Rice and Silverman, 1991). In functional canonical analysis, the relation between a pair of L_2-processes is analyzed by decomposing the processes into their canonical components. The idea of canonical regression analysis is to expand the regression parameter function in terms of functional canonical components for predictor and response processes. The canonical regression decomposition (Theorem 3.3) below provides insights into the structure of the regression parameter functions and not only aids in the understanding of functional linear models, but also leads to a new class of estimation procedures for functional regression analysis. The details of these estimation procedures will be discussed in Section 4, and we demonstrate in Section 5 that these estimates can lead to highly competitive prediction errors in a finite sample situation.

We now state two key results. The first of these (Theorem 3.2) provides the canonical decomposition of the cross-covariance function of processes X and Y. This result plays a central role in the solution of the population normal equation (6). This solution is referred to as canonical regression

decomposition and it leads to an explicit representation of the underlying regression parameter function $\beta_0^*(\cdot,\cdot)$ of the functional linear model (4). The decomposition is in terms of functional canonical correlations ρ_i and canonical weight functions u_i and v_i. Given a predictor process $X(t)$, we obtain as a consequence an explicit representation for $E(Y(t)|X) = (\mathcal{L}_X\beta_0^*)(t)$, where $\mathcal{L}_X$ is as in (4). For the following main results, we refer to the definitions of ρ_i, u_i, v_i, U_i, V_i in Definition 3.1. All proofs are in Section 7.

Theorem 3.2 *(Canonical Decomposition of Cross-covariance Function) Assume L_2-processes X and Y satisfy Condition (C2). Then the cross-covariance function r_{XY} allows the following representation in terms of canonical correlations ρ_i and weight functions u_i and v_i:*

$$r_{XY}(s,t) = \sum_{i=1}^{\infty} \rho_i R_{XX} u_i(s) R_{YY} v_i(t). \tag{15}$$

Theorem 3.3 *(Canonical Regression Decomposition) Assume that the L_2-processes X and Y satisfy Condition (C2). Then one obtains for the regression parameter function $\beta_0^*(\cdot,\cdot)$ (10) the following explicit solution:*

$$\beta_0^*(s,t) = \sum_{i=1}^{\infty} \rho_i u_i(s) R_{YY} v_i(t). \tag{16}$$

To obtain the predicted value of the response process Y, we use the linear predictor

$$E(Y^*(t)|X) = (\mathcal{L}_X\beta_0^*)(t) = \sum_{i=1}^{\infty} \rho_i U_i R_{YY}(v_i)(t). \tag{17}$$

This canonical regression decomposition leads to approximations of the regression parameter function β_0^* and the predicted process $E(Y^*|X) = \mathcal{L}_X\beta_0^*$ via a finitely truncated version of the canonical expansions (16) and (17). The following result provides approximation errors incurred from finite truncation. Thus we have a vehicle to achieve practically feasible estimates of β_0^* and associated predictions Y^* (Section 4).

Theorem 3.4 *For $k \geq 1$, let $\beta_k^*(s\ ,t) = \sum_{i=1}^{k} \rho_i u_i(s) R_{YY}(v_i)(t)$ be the finitely truncated version of the canonical regression decomposition (16) for β_0^*, and define $Y_k^*(t) = (\mathcal{L}_X\beta_k^*)(t)$. Then*

$$Y_k^*(t) = \sum_{i=1}^{k} \rho_i U_i R_{YY}(v_i)(t), \tag{18}$$

with $E[Y_k^*] = 0$. Moreover,

$$E\|Y^* - Y_k^*\|^2 = \sum_{i=k+1}^{\infty} \rho_i^2 \|R_{YY}(v_i)\|^2 \to 0 \text{ as } k \to \infty,$$

and

$$E\|Y - Y_k^*\|^2 = E\|Y\|^2 - E\|\mathcal{L}_X\beta_k^*\|^2 = \text{trace}(R_{YY}) - \sum_{i=1}^{k} \rho_i^2 \|R_{YY}(v_i)\|^2. \tag{19}$$

In finite sample implementations, to be explored in the next two sections, truncation as in (18) is a practical necessity; this requires choice of a suitable truncation parameter k.

4. ESTIMATION PROCEDURES

Estimating the regression parameter function and obtaining fitted processes from the linear model (2) based on a sample of curves is central to the implementation of functional linear models. In practice, data are observed at discrete time points, and we assume for simplicity for the moment that the m_x time points are the same for all observed predictor curves. Likewise, the m_y time points where the response curves are sampled are assumed to be the same. Thus, the original observations are (X_i, Y_i), $i = 1, \ldots, n$, where X_i is an m_x- dimensional vector and Y_i is an m_y- dimensional vector. We assume that m_x and m_y are both large. Without going into any analytical details we compare the finite sample behavior of two functional regression methods, one of which is based on the canonical decomposition and the other on regularization by smoothing.

Our main proposal for a practical functional regression analysis procedure, functional canonical regression analysis (FCR) with local polynomial smoothing, is discussed in Section 4.1. This procedure combines the canonical regression expansion with local polynomial smoothed weight functions. This proposal is contrasted with another procedure which uses the traditional multivariate least squares estimator, coupled with a two-dimensional local polynomial smoothing step (Section 4.2). We refer to this comparison procedure as functional least squares (FLS). For the choice of the smoothing parameters for the various smoothing steps we adopt leave-one-curve-out cross-validation (Rice and Silverman, 1991).

4.1 Functional canonical regression (FCR)

A review of local polynomial smoothing can be found in Fan and Gijbels (1996). Let (U_i, V_i), $i = 1, \ldots, n$, be *i.i.d.* univariate random variables, and (U, V) denote a generic version. To estimate the regression function $m(u) = E(V|U = u)$, a polynomial of order p is fitted by weighted least squares in a local neighborhood around u. The weighted least squares criterion to be minimized is

$$\sum_{i=1}^{n} \{V_i - \sum_{j=0}^{p} \beta_j (U_i - u)^j\}^2 K_h(U_i - u).$$

The resulting curve estimate for $m(x)$ is $\widehat{m}(x) = \widehat{\beta}_0$, where $\widehat{\beta} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_p)'$ is the minimizer for $\beta = (\beta_0, \ldots, \beta_p)'$. Here $K_h(z) \equiv \frac{1}{h} K(\frac{z}{h})$, where $K(\cdot)$ is a nonnegative kernel function (usually with compact support), and h is a bandwidth or smoothing parameter. We write

$$\widehat{m}(u) = S_{LP}(u, (U_i, V_i)_1^n, h).$$

Functional Canonical Regression (FCR) with local polynomial smoothing is then implemented as follows:

(i) In a preprocessing step, all observed process data are centered by subtracting the cross-sectional mean

$$\widetilde{X}_i(s_j) = X_i(s_j) - \frac{1}{m_x}\sum_{l=1}^{m_x} X_l(s_j),$$

and analogously for Y_i. If the data are not sampled on the same grid for different individuals, a smoothing step may be added before the cross-sectional average is obtained. In the following, we use the notation $X(\cdot)$, $X_i(\cdot)$ to denote the centered observations.

(ii) Given a number c of canonical components to be included in FCR, obtain canonical correlations $\breve{\rho}_l$ and weight vectors $(\breve{\mathbf{u}}_l(s_j), j = 1, \ldots, m_x)$, $(\breve{\mathbf{v}}_l(t_k), k = 1, \ldots, m_y)$, $l = 1, \ldots, c$ from the discretized process vectors $(X_i(s_1), \ldots, X_i(s_{m_x}))'$, $(Y_i(t_1), \ldots, Y_i(t_{m_y}))'$, $i = 1, \ldots, n$. This is done by classical multivariate analysis matrix techniques.

(iii) Given a bandwidth h, obtain smoothed (transformed) weight functions as

$$\widetilde{u}_l(s) = S_{LP}(s, (s_j, \breve{\mathbf{u}}_l(s_j))_1^{m_x}, h), \tag{20}$$

$$\widehat{R}_{YY}\widetilde{\mathbf{v}}_l(t) = S_{LP}(t, (t_k, (\breve{\mathbf{v}}_l'\widehat{R}_{YY})(t_k))_1^{m_y}, h),\ l = 1, \ldots, c. \tag{21}$$

Here $\widehat{R}_{YY}$ is an estimate of the discretized sample covariance operator R_{YY}, and is obtained as a $m_y \times m_y$ sample covariance matrix,

$$\widehat{R}_{YY} = \frac{1}{n}\sum_{i=1}^{n} Y_i Y_i'.$$

(iv) Obtain the estimated regression parameter function $\widehat{\beta}$ by applying the canonical regression expansion

$$\widehat{\beta}(s,t) = \sum_{l=1}^{c} \breve{\rho}_l \tilde{u}_l(s)(\hat{R}_{YY}\tilde{v})_l(t).$$

(v) Obtain fitted processes

$$\hat{Y}_i(t) = \int_{T_1} \widehat{\beta}(s,t)\hat{X}_i(s)ds, \quad \text{for} \quad i = 1, \ldots, n,$$

where $\hat{X}_i(s) = S_{LP}(s, (s_j, X_i(s_j))_1^{m_x}, h)$, and the integral is evaluated numerically.

A simplified approximation for step (v), which we adopt in the following and which avoids both smoothing of the observations $X_i(t_j)$ and numerical integration, is

$$\hat{Y}_i(t) = \sum_{l=1}^{m_x} \widehat{\beta}(s_j,t)X_i(s_j)(s_j - s_{j-1}), \tag{22}$$

where we define s_0 so that $s_1 - s_0 = s_2 - s_1$. This procedure depends on two tuning parameters, the bandwidth h for the smoothing steps and the number of canonical components c. These tuning

parameters may be determined by leave-one-out cross-validation (Rice and Silverman, 1991) as follows. With $\alpha = (h, c)$, the leave-one-out estimate for β is

$$\widehat{\beta}_{\alpha}^{(-i)} = \sum_{i=1}^{c} \breve{\rho}_l^{(-i)} \tilde{u}_{h,l}^{(-i)}(s) \hat{R}_{YY}^{(-i)} (\tilde{v})_{h,l}^{(-i)}(t), \quad \text{for} \quad i = 1, \ldots, n, \tag{23}$$

where $\breve{\rho}_l^{(-i)}$ is the l-th canonical correlation, and $\tilde{u}_{h,l}^{(-i)}$ and $\hat{R}_{YY}^{(-i)} \tilde{v}_{h,l}^{(-i)}$ are l-th weight function untransformed and transformed with the covariance operator, respectively. These are obtained as in (20) and (21), using tuning parameter $\alpha = (h, c)$, and omitting the i-th pair of observed curves (X_i, Y_i). The average leave-one-out squared prediction error is then

$$PE_{\alpha} = \frac{1}{n} \sum_{i=1}^{n} \int (Y_i(t) - \int X_i(s) \widehat{\beta}_{\alpha}^{(-i)}(s,t) ds)^2 dt. \tag{24}$$

The cross-validation procedure then selects the tuning parameter that minimizes the approximate average prediction error,

$$\widehat{\alpha} = \arg\min_{\alpha} \widetilde{PE}_{\alpha},$$

where $\widetilde{PE}_{\alpha}$ is obtained by replacing the integrals on the r.h.s. of (24) by sums of the type (22).

4.2. Functional least-squares estimator (FLS)

An alternative procedure is to first obtain a high-dimensional least squares estimator, in the form of a discretized version of the regression parameter function β, which is subsequently regularized by applying a two-dimensional smoothing procedure. While there are many choices regarding the smoothing method, including penalized splines, we apply here local polynomial smoothing due to its simplicity. The resulting functional least squares estimator (FLS) is comparable in spirit to the penalized least squares approach of Ramsay and Dalzell (1991), which is further discussed in Ramsay and Silverman (2002, 2005).

The two-dimensional local polynomial regression smoother we are using is a straightforward extension of the one-dimensional version. For two-dimensional scatterplot data $\{(X_i, Y_i), i = 1, \ldots, n\}$, with $X_i \in \mathbf{R}^2$, $x_i = (X_{i1}, X_{i2})$, $Y_i \in \mathbf{R}$, given a point $x = (x_1, x_2)$, a bandwidth h, and fitting local planes, one minimizes

$$\sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1(X_{i1} - x_1) - \beta_2(X_{i2} - x_2)\}^2 K\left(\frac{X_{i1} - x_1}{h}\right) K\left(\frac{X_{i2} - x_2}{h}\right), \tag{25}$$

with respect to $(\beta_0, \beta_1, \beta_2)$. Then $\hat{m}(x) = \hat{\beta}_0 = S_{2D}(x, (X_i, Y_i)_{i=1}^n, h)$ is the estimate of the regression surface $m(x)$. Other two-dimensional local polynomial regression smoothers could be used, for example employing different bandwidths in different directions, but the simple form (25) suffices for our purposes. The FLS method then consists of the following steps:

(i) The preprocessing step consists in centering all observed processes and is the same as for FCR (see Section 4.1).

(ii) Obtain initial multivariate least squares estimators. Denoting the processes sampled on grids (s_j), $j = 1, \ldots, m_x$, (t_k), $k = 1, \ldots, m_y$ by $\mathbf{X}_i = (X_i(s_j))_1^{m_x}$, $\mathbf{Y}_i = (Y_i(t_k))_1^{m_y}$, $i = 1, \ldots, n$, the multivariate linear model fitted in the first step is $\mathbf{Y} = \mathbf{B}_0\mathbf{X} + \boldsymbol{\varepsilon}$, where $\mathbf{B}_0 = (\beta(s_j, t_k))_{m_x \times m_y}$ is the matrix of regression parameters and $\boldsymbol{\varepsilon} = (\varepsilon_i(s_j))_{n \times m_y}$ are errors. The least squares estimates for $\mathbf{B}_0$ are denoted by $\widetilde{\mathbf{B}}_0 = (\widetilde{b}_{jk})$.

(iii) Smoothing the least squares estimates to obtain a smoothed least squares parameter surface $\beta(s, t)$. This smoothing step is implemented by applying two-dimensional local polynomial smoothing to $\widetilde{\mathbf{B}}_0$,

$$\widehat{\beta}(s, t) = S_{2D}\{(s, t), (s_j, t_k, \widetilde{b}_{jk})_{j=1, k=1}^{m_x, m_y}, h\}.$$

Fitted processes $\widehat{Y}_i$ are obtained as in Step (v) of section 4.1 using approximation (22). Leave-one-out prediction errors are obtained by

$$\widetilde{\mathbf{B}}_0^{(-i)} = (\mathbf{X}^{(-i)\prime}\mathbf{X}^{(-i)})^{-1}\mathbf{X}^{(-i)}\mathbf{Y}^{(-i)},$$

omitting the data for the i-th pair of observed curves. We then obtain $\widehat{\beta}_h^{(-i)}(\cdot, \cdot)$ as in (23), and average approximate prediction errors $\widetilde{PE}_\alpha$ as in (24), where $\alpha = h$. The cross-validation bandwidth choice is then $\widehat{h} = \underset{h}{\operatorname{argmin}} \widetilde{PE}_h$.

5. APPLICATION TO MEDFLY MORTALITY DATA

In this section we present an application to age-at-death data collected for cohorts of male and female medflies. The experiment was carried out to study survival and mortality patterns of cohorts of male and female Mediterranean fruit flies (*Ceratitis capitata*) and is described in Carey et al. (1992). One point of interest is the relation of mortality trajectories between male and female medflies which were raised in the same cage. One desires specifically to quantify the influence of male survival on female survival, as female survival determines the number of eggs laid and thus reproductive success of these flies. We use a subsample of the data generated by this experiment, consisting of 47 cages of medflies, to address these questions. Each cohort consists of approximately 4000 male and 4000 female medflies of the same age which are raised in one shared cage. For each cohort or cage, the number of flies alive at the beginning of each day was recorded; we confined the analysis to the first 40 days of each cohort. The observed processes $X_i(t)$ and $Y_i(t)$, $t = 1, \ldots, 40$, are the estimated random hazard functions for male and female cohorts, respectively. All deaths are fully observed, and censoring is not an issue here.

In order to quantify the relationship between the hazard functions of males and females, we propose to use canonical correlation and weight functions. More importantly, to study the specific influence of male on female mortality for flies that were raised in the same cage, we use a functional linear model, using the hazard function of males as predictor processes and those of females as response

processes. We applied both proposed estimation procedures described in the previous section, with tuning parameters selected by cross-validation.

Table 5.1 lists the average squared errors between fitted and observed processes,

$$SE = \frac{1}{n}\sum_{i=1}^{n}\int_{T_2}[Y_i(t) - \widehat{Y}_i(t)]^2 dt. \tag{26}$$

Here $Y_i(t)$, $\widehat{Y}_i(t)$ are the i-th observed and fitted response processes, respectively. The average squared prediction error (PE) (24) obtained by the one-leave-out technique is also listed. The functional canonical regression (FCR) procedure is seen to perform clearly better than the functional least squares (FLS) procedure with respect to both criteria.

Table 5.1 Results for medfly data, comparing functional canonical regression (FCR) and functional least squares (FLS), regarding average squared prediction error(PE) (24) and squared error (SE) (26). The values for bandwidth h and number c of components chosen by cross-validation are also shown.

	h	c	PE	SE
FCR	10.05	15	0.0069	0.0063
FLS	6.00	NA	0.0167	0.0126

The estimated regression parameter surface $\widehat{\beta}(s,t)$ for FCR using the optimal cross-validation values for h and c (Table 5.1), is shown in Figure 5.1. The regression surface indicates that the effect of male mortality at early age has a large effect on female mortality throughout, probably indicating the effect of the shared environment. At later ages one finds "waves of influence" of the male mortality trajectories on female mortality at increasing ages. Some of these waves quantify negative effects, such as the wave at and before 10 days of male age. Such a negative effect means that increased male mortality in these periods is associated with decreased female mortality. This might be due to various causes, as a lower density of the flies in the cage which is a consequence of increased male mortality may be associated with increased survival; or fewer males may lead to lower "cost of mating" for females, with associated lower mortality.

Some examples of observed male and female as well as fitted and predicted female mortality trajectories for three randomly selected pairs of cohorts (cages) are displayed in Figure 5.2. The predicted female trajectories were constructed by applying the respective regression method with the leave-one-out technique. We find that FCR provides seemingly better predictions in this application. We note the presence of a "shoulder" at around day 20 for the three female mortality curves. This mortality shoulder has been previously found and characterized as an expression of the cost of reproduction for female medflies (see Müller, et al., 1997). The male medflies do not have such a prominent shoulder, but rather exhibit a mortality peak at a much later time around day 30. It is of interest that the functional regression method correctly predicts the shoulder effect in female mortality. At the rightmost points, for ages near 40 days, the variability of the mortality trajectories becomes so large that prediction of the right tails of the trajectories is difficult.

6. SOME ADDITIONAL RESULTS

Theorems 6.3 and 6.4 in this section provide a functional analogue to the sums of squares decomposition of classical regression analysis. In addition, we provide two results characterizing the regression operators $\mathcal{L}_X$. We begin with two auxiliary results which are taken from He et al. (2003). The first of these characterizes the correlation operator between processes X and Y.

Lemma 6.1 *Assume the L_2-processes X and Y satisfy Condition (C2). Then the correlation operator $R_{XX}^{-1/2}R_{XY}R_{YY}^{-1/2}$ can be extended continuously to a Hilbert-Schmidt operator R on $L_2(T_2)$ to $L_2(T_1)$. Hence $R_0 = R^*R$ is also a Hilbert-Schmidt operator with a countable number of non-zero eigenvalues and eigenfunctions $\{(\lambda_i, q_i)\}$, $i \geq 1, \lambda_1 \geq \lambda_2 \geq \ldots,\ p_i = Rq_i/\sqrt{\lambda_i}$. Then*

(a) $\rho_i = \sqrt{\lambda_i}$, $u_i = R_{XX}^{-1/2}p_i$, $v_i = R_{YY}^{-1/2}q_i$ *and both u_i and v_i are L_2-functions;*

(b) $\mathrm{corr}(U_i, U_i) = \langle u_i, R_{XX}u_j\rangle = \langle p_i, p_j\rangle = \delta_{ij}$;

(c) $\mathrm{corr}(V_i, V_i) = \langle v_i, R_{XX}v_j\rangle = \langle q_i, q_j\rangle = \delta_{ij}$;

(d) $\mathrm{corr}(U_i, V_j) = \langle u_i, R_{XX}v_j\rangle = \langle p_i, Rq_j\rangle = \rho_i\delta_{ij}$.

One of the main results in He et al. (2003) reveals that the L_2-processes X and Y can be expressed as sums of uncorrelated component functions and the correlation between the ith components of the expansion is the ith corresponding functional canonical correlation between the two processes.

Lemma 6.2 *(Canonical Decomposition) Assume L_2-processes X and Y satisfy Condition (C2). Then there exists a decomposition*

(a) $$X = X_{c,k} + X_{c,k}^{\perp},\ Y = Y_{c,k} + Y_{c,k}^{\perp},$$

where

$$X_{c,k} = \sum_{i=1}^{k} U_i R_{XX} u_i, \quad X_{c,k}^{\perp} = X - X_{c,k}, \quad Y_{c,k} = \sum_{i=1}^{k} V_i R_{YY} v_i, \quad Y_{c,k}^{\perp} = Y - Y_{c,k}.$$

The indices c, k stand for canonical decomposition with k components, and U_i, V_i, u_i, v_i, are as in Definition 3.1. Here (X, Y) and $(X_{c,k}, Y_{c,k})$ share the same first k canonical components; and $(X_{c,k}, Y_{c,k})$ and $(X_{c,k}^{\perp}, Y_{c,k}^{\perp})$ are uncorrelated, i.e.,

$$\mathrm{corr}(X_{c,k}, X_{c,k}^{\perp}) = 0, \quad \mathrm{corr}(Y_{c,k}, Y_{c,k}^{\perp}) = 0, \quad \mathrm{corr}(X_{c,k}, Y_{c,k}^{\perp}) = 0, \quad \mathrm{corr}(Y_{c,k}, X_{c,k}^{\perp}) = 0.$$

(b) Let $k \to \infty$, and $X_{c,\infty} = \sum_{i=1}^{\infty} U_i R_{XX} u_i$, $\quad Y_{c,\infty} = \sum_{i=1}^{\infty} V_i R_{YY} v_i$. Then

$$X = X_{c,\infty} + X_{c,\infty}^{\perp},\ \ Y = Y_{c,\infty} + X_{c,\infty}^{\perp},$$

where $X_{c,\infty}^{\perp} = X - X_{c,\infty}$, $Y_{c,\infty}^{\perp} = Y - Y_{c,\infty}$. Here, $(X_{c,\infty}, Y_{c,\infty})$ and (X, Y) share the same canonical components, $\mathrm{corr}(X_{c,\infty}^{\perp}, Y_{c,\infty}^{\perp}) = 0$, and $(X_{c,\infty}^{\perp}, Y_{c,\infty}^{\perp})$ and $(X_{c,\infty}, Y_{c,\infty})$ are uncorrelated. Moreover, $X_{c,\infty}^{\perp} = 0$ if $\{p_i, i \geq 1\}$ forms a basis of the closure of the domain of R_{XX}, and $Y_{c,\infty}^{\perp} = 0$ if $\{q_i, i \geq 1\}$ forms a basis of the closure of the domain of R_{YY}.

Since the covariance operators of L_2-processes are nonnegative self-adjoint, they can be ordered as follows. The definitions of Y^*, Y_k^*, $Y_{c,\infty}$ are in Theorems 3.4, 3.5 and Lemma 6.2(b), respectively.

Theorem 6.3 *For $k \geq 1$, $R_{Y_k^* Y_k^*} \leq R_{Y^* Y^*} \leq R_{Y_{c,\infty} Y_{c,\infty}} \leq R_{YY}$.*

In multiple regression analysis, the ordering of the operators in Theorem 6.3 is related to the ordering of regression models in terms of a notion analogous to the regression sum of squares (SSR). The canonical regression decomposition provides information about the model in terms of its canonical components. Our next result describes the canonical correlations between observed and fitted processes. This provides an extension of the coefficient of multiple determination, $R^2 = \mathrm{corr}(Y, \widehat{Y})$, an important quantity in classical multiple regression analysis, to the functional case; compare also Yao et al. (2005).

Theorem 6.4 *Assume that L_2-processes X and Y satisfy Condition (C2). Then the canonical correlations and weight functions for the pair of obsered and fitted response processes (Y, Y^*) are $\{(\rho_i, v_i, v_i/\rho_i); i \geq 1\}$, and the corresponding k-component (or ∞-component) canonical decomposition for Y^*, as defined in Lemma 6.2 and denoted here by $Y_{c,k}^*$(or $Y_{c,\infty}^*$) is equivalent to the process Y_k^* or Y^* given in Theorem 3.4, i.e.,*

$$Y_{c,k}^* = Y_k^* - \sum_{i=1}^{k} \rho_i U_i R_{YY} v_i, \ k \geq 1, \quad Y_{c,\infty}^* = Y^* - \sum_{i=1}^{k} \rho_i U_i R_{YY} v_i. \tag{27}$$

We note that if Y is a scalar, then $R^2 = \rho_1$, and for a functional Y, R^2 is replaced by the set $\{\rho_i, i \geq 1\}$.

The following two results serve to characterize the regression operator $\mathcal{L}_X$ defined in (4). They are used in the proofs provided in the following section.

Proposition 6.5 *The adjoint operator of $\mathcal{L}_X$ is $\mathcal{L}_X^* : L_2(T_2) \to L_2(T_1 \times T_2)$, where*

$$(\mathcal{L}_X^* z)(s,t) = X(s) z(t), \ for \ z \in L_2(T_2).$$

We have the following relation between correlation operator Γ_{XX} defined in (5) and the regression operator $\mathcal{L}_X$.

Proposition 6.6 *The operator Γ_{XX} is a self-adjoint, non-negative Hilbert-Schmidt operator and satisfies $\Gamma_{XX} = E[\mathcal{L}_X^* \mathcal{L}_X]$.*

7. PROOFS

In this section, we provide sketches of proofs and some auxiliary results. We use the tensor notation to define an operator $\theta \otimes \varphi : H \to H$,

$$(\theta \otimes \varphi)(h) = \langle h, \theta \rangle \varphi, \ for \ h \in H.$$

Proof of Proposition 2.2. To prove (a) $\Rightarrow$ (b), we multiply equation (4) with X on both sides and take expected values to obtain $E(XY) = E(X\mathcal{L}_X \beta_0) + E(X\varepsilon)$. Then equation (6) follows from $E(XY) = r_{XY}$, $E(X\mathcal{L}_X \beta_0) = \Gamma_{XX}\beta_0$ (by Propositions 6.5 and 6.6), and $E(X\varepsilon) = 0$.

For (b)$\Rightarrow$ (c), let β_0 be a solution of equation (6). Then for any $\beta \in L_2(T_1 \times T_2)$, $E\|Y - \mathcal{L}_X\beta\|^2 = E\|Y - \mathcal{L}_X\beta_0\|^2 + E\|\mathcal{L}_X(\beta_0 - \beta)\|^2 + 2E[\langle Y - \mathcal{L}_X\beta_0, \mathcal{L}_X(\beta_0 - \beta)\rangle]$. Since

$$\begin{aligned} E\langle Y &- \mathcal{L}_X\beta_0, \mathcal{L}_X(\beta_0 - \beta)\rangle = E\langle \mathcal{L}_X^* Y - \mathcal{L}_X^*\mathcal{L}_X\beta_0, \beta_0 - \beta\rangle \\ &= \langle E[\mathcal{L}_X^* Y] - E[\mathcal{L}_X^*\mathcal{L}_X\beta_0], \beta_0 - \beta\rangle = \langle r_{XY} - \Gamma_{XX}\beta_0, \beta_0 - \beta\rangle = 0, \end{aligned}$$

by Proposition 6.6, we then have

$$E\|Y - \mathcal{L}_X\beta\|^2 = E\|Y - \mathcal{L}_X\beta_0\|^2 + E\|\mathcal{L}_X(\beta_0 - \beta)\|^2 \geq E\|Y - \mathcal{L}_X\beta_0\|^2,$$

which implies that β_0 is indeed a minimizer of $E\|Y - \mathcal{L}_X\beta\|^2$.

For (c)$\Rightarrow$(a), let

$$d^2 = E\|Y - \mathcal{L}_X\beta_0\|^2 = \min_{\beta \in L_2(T_1 \times T_2)} E\|Y - \mathcal{L}_X\beta\|^2.$$

Then, for any $\beta \in L_2(T_1 \times T_2)$, $a \in \mathbf{R}$,

$$\begin{aligned} d^2 &= E\|Y - \mathcal{L}_X\beta_0\|^2 \leq E\|Y - \mathcal{L}_X(\beta_0 + a\beta)\|^2 \\ &= E\|Y - \mathcal{L}_X\beta_0\|^2 - 2E\langle Y - \mathcal{L}_X\beta_0, \mathcal{L}_X(a\beta)\rangle + E\|\mathcal{L}_X(a\beta)\|^2 \\ &= d^2 - 2a\langle E[X(Y - \mathcal{L}_X\beta_0)], \beta\rangle + a^2 E\|\mathcal{L}_X\beta\|^2. \end{aligned}$$

Choosing $a = \langle E[X(Y - \mathcal{L}_X\beta_0)], \beta\rangle / E\|\mathcal{L}_X\beta\|^2$, it follows that $|\langle E[X(Y - \mathcal{L}_X\beta_0)], \beta\rangle|^2 / E\|\mathcal{L}_X\beta\|^2 \leq 0$ and $\langle E[X(Y - \mathcal{L}_X\beta_0)], \beta\rangle = 0$. Since β is arbitrary, $E[X(Y - \mathcal{L}_X\beta_0)] = 0$, and therefore β_0 satisfies the functional linear model (4). $\Diamond$

Proof of Theorem 2.3. Note first that $r_{XY}(s,t) = \sum_{i,j} E[\xi_i\zeta_j]\theta_i(s)\varphi_j(t)$. Thus Condition (C1) is equivalent to $r_{XY} \in G_{XX}$. Suppose a unique solution of (4) exists in $\ker(\Gamma_{\mathrm{XX}})^\perp$. Then this solution is also a solution of (6) by Theorem 2.2(b). Therefore, $r_{XY} \in G_{XX}$, which implies (C1). On the other hand, if (C1) holds, then $r_{XY} \in G_{XX}$, which implies $\Gamma_{XX}^{-1} r_{XY} = \sum_i \lambda_{Xi}^{-1}\langle r_{XY}, \theta_i\varphi_j\rangle\theta_i\varphi_j$ is a solution of (6) and is in $\ker(\Gamma_{\mathrm{XX}})^\perp$, and therefore is the unique solution in $\ker(\Gamma_{\mathrm{XX}})^\perp$, and also the unique solution of (4) in $\ker(\Gamma_{\mathrm{XX}})^\perp$. $\Diamond$

Proof of Proposition 2.4 The equivalence of (a), (b) and (c) follows from Proposition 2.2, and (d) $\Rightarrow$ (b) is a consequence of Proposition 2.3. We now prove (b) $\Rightarrow$ (d). Let β_0 be a solution of (6). Propositions 2.2 and 2.3 imply that both β_0 and β_0^* minimize $E\|Y - \mathcal{L}_X\beta\|^2$, for $\beta \in L_2(T_1 \times T_2)$. Hence $E\|Y - \mathcal{L}_X\beta_0\|^2 = E\|Y - \mathcal{L}_X\beta_0^*\|^2 + E\|\mathcal{L}_X(\beta_0^* - \beta_0)\|^2 + 2E\langle Y - \mathcal{L}_X\beta_0^*, \mathcal{L}_X(\beta_0^* - \beta_0)\rangle$, which by Proposition 6.6 implies $2E\langle \mathcal{L}_X^*(Y - \mathcal{L}_X\beta_0^*), \beta_0^* - \beta_0\rangle = 2\langle r_{XY} - \Gamma_{XX}\beta_0^*, \beta_0^* - \beta_0\rangle = 0$. Therefore, $E\|\mathcal{L}_X(\beta_0^* - \beta_0)\|^2 = \|\Gamma_{XX}^{1/2}(\beta_0^* - \beta_0)\|^2 = 0$. This implies $\beta_0^* - \beta_0 \in \ker(\Gamma_{\mathrm{XX}})$, or $\beta_0 = \beta_0^* + h$, for an $h \in \ker(\Gamma_{\mathrm{XX}})$. $\Diamond$

Proof of Theorem 3.2. According to Lemma 6.2 (b), Condition (C2) guarantees the existence of

the canonical components and canonical decomposition of X and Y. Moreover,

$$\begin{aligned} r_{XY}(s,t) &= E[X(s)Y(t)] = E[(X_{c,\infty}(s) + X^{\perp}_{c,\infty}(s))(Y_{c,\infty}(t) + Y^{\perp}_{c,\infty}(t))] \\ &= E[X_{c,\infty}(s)Y_{c,\infty}(t)] = E[\sum_{i=1}^{\infty} U_i R_{XX} u_i(s) \sum_{i=1}^{\infty} V_i R_{YY} v_i(t)] \\ &= \sum_{i,j=1}^{\infty} E[U_i V_j] R_{XX} u_i(s) R_{YY} v_i(t) = \sum_{i=1}^{\infty} \rho_i R_{XX} u_i(s) R_{YY} v_i(t). \end{aligned}$$

We now show that the exchange of the expectation with the summation above is valid. From Lemma 6.1 (b) for any $k > 0$, and the spectral decomposition $R_{XX} = \sum_i \lambda_{Xi} \theta_i \otimes \theta_i$,

$$\begin{aligned} \sum_{i=1}^{k} E\|U_i R_{XX} u_i\|^2 &= \sum_{i=1}^{k} E[U_i^2] \|R_{XX}^{1/2} p_i\|^2 = \sum_{i=1}^{k} \langle p_i, R_{XX} p_i \rangle \\ &= \sum_{i=1}^{k} \sum_{j=1}^{\infty} \lambda_{Xj} \langle p_i, \theta_j \rangle^2 = \sum_{j=1}^{\infty} \lambda_{Xj} (\sum_{i=1}^{k} \langle p_i, \theta_j \rangle^2) \\ &\leq \sum_{j=1}^{\infty} \lambda_{Xj} \|\theta_j\|^2 = \sum_{j=1}^{\infty} \lambda_{Xj} < \infty, \end{aligned}$$

where the inequality follows from the fact that $\sum_{i=1}^{k} \langle p_i, \theta_j \rangle^2$ is the square length of the projection of θ_j onto the linear subspace spanned by $\{p_1, \ldots, p_k\}$. Similarly, we can show that for any $k > 0$,

$$\sum_{i=1}^{k} E\|V_i R_{YY} v_i\|^2 < \sum_{j=1}^{\infty} \lambda_{Yj} < \infty. \qquad \Diamond$$

Proof of Theorem 3.3. Note that Condition (C2) implies Condition (C1). Hence, from Theorem 2.3, $\beta_0^* = \Gamma_{XX}^{-1} r_{XY}$ exists and is unique in $\ker(\Gamma_{\mathrm{XX}})^{\perp}$. We can show (16) by applying Γ_{XX}^{-1} to both sides of (6), exchanging the order of summation and integration. To establish (17), it remains to show

$$\sum_{i=1}^{\infty} \|\rho_i u_i R_{YY} v_i\|^2 < \infty, \tag{28}$$

where $u_i R_{YY} v_i(s,t) = u_i(s) R_{YY} v_i(t)$ in $L_2(T_1 \times T_2)$. Note that

$$\rho_i u_i = \rho_i R_{XX}^{1/2} p_i = R_{XX}^{1/2} R q_i = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_{Xj}}} \langle R q_i, \theta_j \rangle \theta_j,$$

where the operator $R = R_{XX}^{1/2} R_{XY} R_{YY}^{1/2}$ is defined in Lemma 6.1 and can be written as $R = \sum_{k,m} r_{km} \varphi_k \otimes \theta_m$, with $r_{km} = E[\xi_k \zeta_m]/\sqrt{\lambda_{Xk} \lambda_{Ym}}$, using the Karhunen -Loève expansion (7). Then

$$R q_i = \sum_{k,m} r_{km} \langle \varphi_k, q_i \rangle \theta_m, \quad \langle R q_i, \theta_j \rangle = \sum_{k} r_{kj} \langle \varphi_k, q_i \rangle.$$

Therefore,

$$\sum_i \|\rho_i u_i R_{YY} v_i\|^2 \le \sum_i \|\rho_i u_i\|^2 \|R_{YY} v_i\|^2 = \sum_i [\sum_j \frac{1}{\lambda_{Xj}} (\langle R q_i, \theta_j \rangle)^2] \, \|R_{YY} v_i\|^2$$

$$= \sum_i [\sum_j \frac{1}{\lambda_{Xj}} \left\{ \sum_k r_{kj} \langle \varphi_k, q_i \rangle \right\}^2] \|R_{YY} v_i\|^2 \le \sum_i [\sum_j \frac{1}{\lambda_{Xj}} \sum_k r_{kj}^2 \sum_m \langle \varphi_m, q_i \rangle^2] \|R_{YY} v_i\|^2$$

$$= [\sum_j \frac{1}{\lambda_{Xj}} \sum_k r_{kj}^2] \sum_i [\sum_m \langle \varphi_m, q_i \rangle^2] \|R_{YY} v_i\|^2 = \sum_{j,k} \frac{r_{kj}^2}{\lambda_{Xj}} \sum_i \|R_{YY} v_i\|^2, \qquad \text{as } \|q_i\| = 1.$$

Note that by (C2) the first sum on the r.h.s. is bounded. For the second sum,

$$\sum_i \|R_{YY} v_i\|^2 = \sum_i \|R_{YY}^{1/2} q_i\|^2 = \sum_i \langle q_i, R_{YY} q_i \rangle = \sum_i \sum_j \lambda_{Yj} \langle q_i, \varphi_j \rangle^2$$

$$= \sum_j \lambda_{Yj} \sum_i \langle q_i, \varphi_j \rangle^2 \le \sum_j \lambda_{Yj} \|\varphi_j\|^2 \le \sum_j \lambda_{Yj} < \infty,$$

which implies (28). $\Diamond$

Proof of Theorem 3.4. Observing

$$\begin{aligned}
Y_k^* = \mathcal{L}_X \ \beta_k^* &= \sum_{i=1}^k \rho_i \mathcal{L}_X(u_i) R_{YY}(v_i) = \sum_{i=1}^k \rho_i \langle u_i, X \rangle R_{YY}(v_i) = \sum_{i=1}^k \rho_i U_i R_{YY}(v_i), \\
E\|Y^* - Y_k^*\| &= E\| \sum_{i=k+1}^\infty \rho_i U_i R_{YY}(v_i)\|^2 = \sum_{i=k+1}^\infty \rho_i \|R_{YY}(v_i)\|^2 \quad \text{and} \\
E\|\mathcal{L}_X \beta_k^*\|^2 &= E\| \sum_{i=1}^\infty \rho_i U_i R_{YY}(v_i)\|^2 \\
&= \sum_{i,j=1}^\infty \rho_i \rho_j E[U_i U_j] \langle R_{YY}(v_i), R_{YY}(v_j) \rangle = \sum_{i=1}^\infty \rho_i^2 \|R_{YY}(v_i)\|^2 < \infty,
\end{aligned}$$

we infer $E\|Y^* - Y_k^*\|^2 \to 0$, as $k \to \infty$. From $E[U_i] = 0$, for $i \ge 1$, we have $E[Y_k^*] = 0$, and moreover

$$\begin{aligned}
& E\|Y - Y_k^*\|^2 = E\|(Y - \mathcal{L}_X \ \beta_0^*) + \mathcal{L}_X(\beta_0^* - \beta_k^*)\|^2 \\
= \ & E\|Y - \mathcal{L}_X \ \beta_0^*\|^2 + E\|\mathcal{L}_X(\beta_0^* - \beta_k^*)\|^2 + 2E\langle Y - \mathcal{L}_X \ \beta_0^*, \mathcal{L}_X(\beta_0^* - \beta_k^*) \rangle.
\end{aligned}$$

Since $E\|Y - \mathcal{L}_X \ \beta_0^*\|^2 = \text{trace}(R_{YY}) - E\|\mathcal{L}_X \beta_0^*\|^2$, and as β_0^* is the solution of the normal equation (6), we obtain $E\langle Y - \mathcal{L}_X \ \beta_0^*, \mathcal{L}_X(\beta_0^* - \beta_k^*) \rangle = E\langle \mathcal{L}_X^*(Y - \mathcal{L}_X \ \beta_0^*), \beta_0^* - \beta_k^* \rangle = 0$. Likewise,

$$E\|\mathcal{L}_X(\beta_0^* - \beta_k^*)\|^2 = \sum_{i=k+1}^\infty \rho_i^2 \|R_{YY}(v_i)\|^2,$$

and (19) follows. $\Diamond$

Proof of Theorem 6.3. From (17), (18) for any $k \ge 1$,

$$R_{Y^*Y^*} - R_{Y_k^* Y_k^*} = R_{YY}^{1/2} [\sum_{i=k+1}^\infty \rho_i^2 q_i \otimes q_i] \, R_{YY}^{1/2} = R_{YY}^{1/2} R_{k+1}^* R_{k+1} R_{YY}^{1/2},$$

where $R_{k+1} = \mathrm{Proj}_{\mathrm{span}\{q_i, i \geq k+1\}} R$, and hence, $R_{Y^*Y^*} - R_{Y_k^* Y_k^*} \geq 0$. Note that

$$\begin{aligned} r_{Y_{c,\infty}Y_{c,\infty}}(s,t) &= E[Y_{c,\infty}(s)Y_{c,\infty}(t)] = \sum_{i,j=1}^{\infty} E[V_i V_j] R_{YY}(v_i)(s) R_{YY}(v_j)(t) \\ &= \sum_{i=1}^{\infty} R_{YY}(v_i)(s) R_{YY}(v_j)(t) = \sum_{i=1}^{\infty} R_{YY}^{1/2}(q_i)(s) R_{YY}^{1/2}(q_i)(t), \end{aligned}$$

implying

$$R_{Y_{c,\infty}Y_{c,\infty}} - R_{Y^*Y^*} = R_{YY}^{1/2}[\sum_{i=1}^{\infty}(1-\rho_i^2) q_i \otimes q_i] R_{YY}^{1/2} \geq 0.$$

Finally, from Lemma 6.2(b), we have $Y = Y_{c,\infty} - Y_{c,\infty}^{\perp}$, which implies $r_{YY} = r_{Y_{c,\infty}Y_{c,\infty}} + r_{Y_{c,\infty}^{\perp}Y_{c,\infty}^{\perp}}$. Therefore, $r_{YY} - r_{Y_{c,\infty}Y_{c,\infty}} = r_{Y_{c,\infty}^{\perp}Y_{c,\infty}^{\perp}}$, and $R_{YY} - R_{Y_{c,\infty}Y_{c,\infty}} = R_{Y_{c,\infty}^{\perp}Y_{c,\infty}^{\perp}} \geq 0$. $\Diamond$

We need the following auxiliary result to prove Theorem 6.3. We call two L_2-processes X and Y uncorrelated if and only if $E[\langle u, X\rangle \langle v, Y\rangle] = 0$ for all L_2-functions u and v.

Lemma 7.1 *$Y_{c,\infty}^{\perp}$ and Y^* are uncorrelated.*

Proof. For any $\widetilde{u}, \widetilde{v} \in L_2(T_2)$, write $\widetilde{v} = \widetilde{v}_1 + \widetilde{v}_2$, with $R_{YY}^{1/2}\, \widetilde{v}_1 \in \mathrm{span}\{q_i; \quad i \geq 1\}$ which is equivalent to $\widetilde{v}_1 \in \mathrm{span}\{v_i; \quad i \geq 1\}$, and $R_{YY}^{1/2}\, \widetilde{v}_2 \in \mathrm{span}\{q_i; \quad i \geq 1\}^{\perp}$. Then,

$$\langle \widetilde{v}_2, Y^* \rangle = \sum_{i=1}^{\infty} \rho_i U_i \langle \widetilde{v}_2, R_{YY} v_i \rangle = \sum_{i=1}^{\infty} \rho_i U_i \langle R_{YY}^{1/2} \widetilde{v}_2, q_i \rangle = 0.$$

With $\widetilde{v}_1 = \sum_i a_i v_i$, $\quad \langle \widetilde{v}, Y^* \rangle = \langle \widetilde{v}_1, Y^* \rangle = \sum_{i,j} a_i \rho_j U_j \langle v_i, R_{YY} v_j \rangle = \sum_i a_i \rho_i U_i$. Furthermore, from Lemma 6.2 (b), $E[U_i \langle \widetilde{u}, Y_{c,\infty}^{\perp} \rangle] = 0$ for all $i \geq 1$, and we have $E[\langle \widetilde{u}, Y_{c,\infty}^{\perp} \rangle \langle \widetilde{v}, Y^* \rangle] = 0$. $\Diamond$

Proof of Theorem 6.4. Calculating the covariance operators for (Y, Y^*),

$$\begin{aligned} r_{Y^*Y^*}(s,t) &= E[Y^*(s)Y^*(t)] = \sum_{i,j} \rho_i \rho_j E[U_i U_j] R_{YY} u_i(s) R_{YY} v_j(t) \\ &= \sum_i \rho_i^2 R_{YY} u_i(s) R_{YY} v_i(t) = \sum_i \rho_i^2 R_{YY}^{1/2} q_i(s) R_{YY}^{1/2} q_i(t), \end{aligned}$$

so that

$$R_{Y^*Y^*} = \sum_i \rho_i^2 R_{YY}^{1/2}[q_i \otimes q_i] R_{YY}^{1/2} = R_{YY}^{1/2}[\sum_i \rho_i^2 q_i \otimes q_i] R_{YY}^{1/2} = R_{YY}^{1/2} R_0 R_{YY}^{1/2}.$$

Now from Lemmas 6.2 and 7.1,

$$\begin{aligned} r_{YY^*}(s,t) &= E[Y(s)Y^*(t)] = E[(Y_{c,\infty}(s) + Y_{c,\infty}^{\perp}(s))Y^*(t)] \\ &= E[Y_{c,\infty}(s)Y^*(t)] = E[\sum_i V_i R_{YY} v_i(s) \sum_j \rho_j U_j R_{YY} v_j(t) \\ &= \sum_{i,j} E[V_i U_j \rho_j R_{YY} v_i(s) R_{YY} v_j(t)] \\ &= \sum_i \rho_i^2 R_{YY} v_i(s) R_{YY} v_j(t) = r_{Y^*Y^*}(s,t). \end{aligned}$$

Hence, $R_{YY^*} = R_{Y^*Y^*}$. The correlation operator for(Y, Y^*) is $\widetilde{R} = R_{YY}^{-1/2} R_{YY^*} R_{Y^*Y^*}^{-1/2} = R_{YY}^{-1/2} R_{Y^*Y^*}^{1/2}$, with $\widetilde{R}\widetilde{R}^* = R_{YY}^{-1/2} R_{Y^*Y^*} R_{YY}^{-1/2} = R_0$. Hence, $\widetilde{\rho}_i = \rho_i$, $\widetilde{p}_i = q_i$, and $\widetilde{q}_i = \widetilde{R}^* \widetilde{p}_i / \widetilde{\rho}_i = R_{Y^*Y^*}^{1/2} R_{YY}^{-1/2} q_i / \rho_i = R_{Y^*Y^*}^{1/2} v_i / \rho_i$. Moreover, $\widetilde{u}_i = R_{YY}^{-1/2} \widetilde{p}_i = R_{YY}^{-1/2} q_i = v_i$, $\quad \widetilde{v}_i = R_{Y^*Y^*}^{-1/2} \widetilde{q}_i = R_{Y^*Y^*}^{-1/2} R_{Y^*Y^*}^{1/2} v_i / \rho_i = v_i / \rho_i$. Note $Y_{c,\infty}^* = \sum_i \widetilde{V}_i R_{Y^*Y^*} \widetilde{v}_i$, with

$$\widetilde{V}_i = \langle \widetilde{v}_i, Y^* \rangle = \langle v_i / \rho_i, \sum_j \rho_j U_j R_{YY} v_j \rangle = \sum_j U_j \langle v_i, R_{YY} v_j \rangle = U_i,$$

$$R_{Y^*Y^*} \widetilde{v}_i = R_{YY}^{1/2} R_0 R_{YY}^{1/2} v_i / \rho_i = R_{YY}^{1/2} R_0 q_i / \rho_i = \rho_i R_{YY}^{1/2} q_i = \rho_i R_{YY} v_i.$$

Substituting into the left equation of (27), one obtains the right equation of (27). $\Diamond$

Proof of Proposition 6.5. From the definition, $\mathcal{L}_X^*$ must satisfy $\langle \mathcal{L}_X \beta,\, z \rangle = \langle \beta, \mathcal{L}_X^* z \rangle$, for $\beta \in L_2(T_1 \times T_2)$, and $z \in L_2(T_2)$. Note that $\langle \mathcal{L}_X \beta,\, z \rangle = \int_{T_2} (\mathcal{L}_X \beta)(t) z(t) dt = \int_{T_2} \int_{T_1} X(s) \beta(s,t) z(t) ds dt$ and $\langle \beta, \mathcal{L}_X^* z \rangle = \int \int_{T_1 \times T_2} \beta(s,t) (\mathcal{L}_X^* z)(s,t) ds dt$. Then, from the difference of the two, we have $\int \int \beta(s,t) [X(s) z(t) - (\mathcal{L}_X^* z)(s,t)] ds dt = 0$ for arbitrary $\beta \in L_2(T_1 \times T_2)$ and $z \in L_2(T_2)$. This implies that $(\mathcal{L}_X^* z)(s,t) = X(s) z(t)$. $\Diamond$

Proof of Proposition 6.6. By Proposition 6.5, $\Gamma_{XX} = E[\mathcal{L}_X^* \mathcal{L}_X]$. Since the integral operator Γ_{XX} has the L_2 integral kernel r_{XX}, it is a Hilbert-Schmidt operator (Conway, 1985). Moreover, for $\beta_1, \beta_2 \in L_2(T_1 \times T_2)$,

$$\begin{aligned} \langle \Gamma_{XX} \beta_1, \beta_2 \rangle &= \int \int (\Gamma_{XX} \beta_1)(s,t) \beta_2(s,t) ds dt = \int \int \int r_{XX}(s,w) \beta_1(w,t) \beta_2(s,t) dw ds dt, \\ \langle \beta_1, \Gamma_{XX} \beta_2 \rangle &= \int \int \beta_1(s,t) (\Gamma_{XX} \beta_2(s,t)) ds dt = \int \int \int \beta_1(w,t) r_{XX}(s,w) \beta_2(s,t) dw ds dt, \end{aligned}$$

implying that Γ_{XX} is self-adjoint. Furthermore, Γ_{XX} is non-negative definite, because for arbitrary $\beta \in L_2(T_1 \times T_2)$,

$$\begin{aligned} \langle \Gamma_{XX} \beta, \beta \rangle &= \int \int \int E[X(s) X(w)] \beta(w,t) \beta(s,t) dw ds dt \\ &= E[\int (\mathcal{L}_X \beta)(t) (\mathcal{L}_X \beta)(t) dt] = E \| \mathcal{L}_X \beta \|^2 \geq 0. \qquad \Diamond \end{aligned}$$

REFERENCES

Ash, R. B. and Gardner, M. F. (1975), Topics in Stochastic Processes, New York: Academic Press.

Brillinger, D. R. (1985), Time Series: Data Analysis and Theory, Second Edition. Holden Day, San Francisco.

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

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

Conway, J. B. (1985). *A Course in Functional Analysis.* New York: Springer-Verlag.

Eubank, R. and Hsing, T. (2007). Canonical correlation for stochastic processes. Preprint.

Fan, J. and Zhang, J. T. (1998). Functional linear models for longitudinal data. *J. Royal Statist. Assoc. Series B* **39**, 254-261.

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

Hall, P. and Horowitz, J. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70-91.

Hannan, E. J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J. Australian Math. Soc.* **2**, 229242.

He, G., Müller, H.G. and Wang, J.L. (2003). Functional canonical analysis for square integrable stochastic processes. *J. Multiv. Anal.* **85**, 54-77.

He, G., Müller, H.G., Wang, J.L. (2004). Methods of canonical analysis for functional data. *J. Statist. Plann. and Inf.* **122**, 141-159.

Hoover, D.R., Rice, J., Wu, C.O. and Yang, L. (1998). Nonparametric smoothing estimates of time varying coefficient models with longitudinal data. *Biometrika* **85**, 809-740.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* **28**, 321-377.

Leurgans, S.E., Moyeed, R.A. and Silverman, B.W. (1993). Canonical correlation analysis when the data are curves. *J. Royal Statist. Soc. Series B* **55**, 725-740.

Müller, H.G., Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics* **33**, 774-805.

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

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

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

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

Shi, M., Weiss, R.E. and 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., Müller, H.G. and Wang, J.L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics* **33**, 2873-2903.

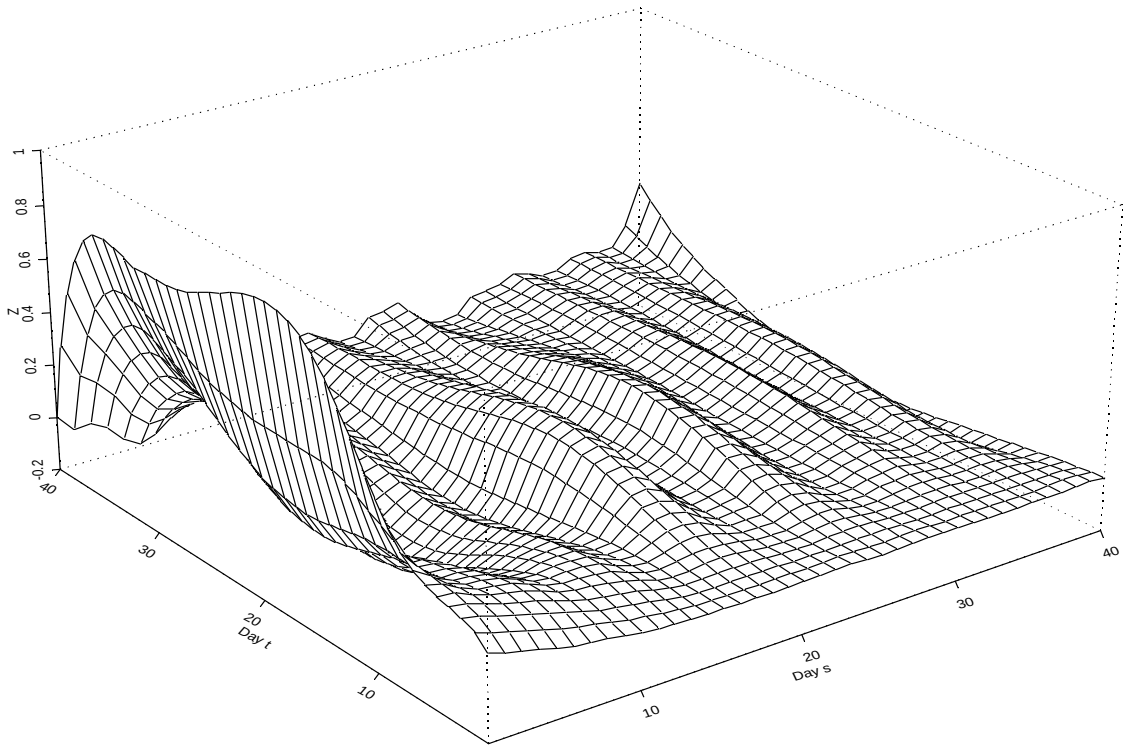

Figure 1: Estimated regression parameter surface via functional canonical regression for the medfly study.

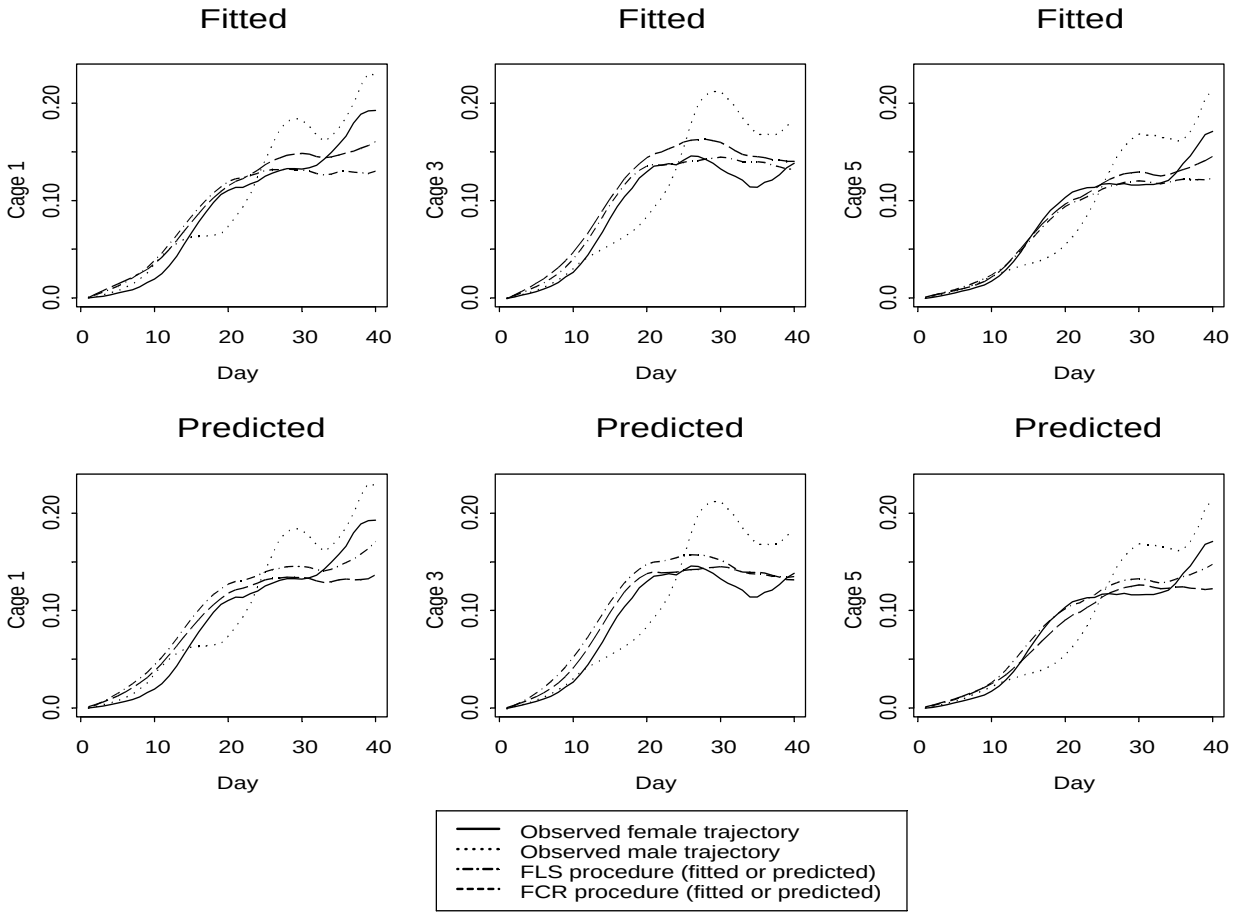

Figure 2: Functional regression of female (response) on male (predictor) medfly trajectories of mortality, data for three cages. Shown are observed male and female trajectories as well as fitted and predicted female trajectories, using estimation procedures based on functional last squares and on functional canonical regression.