

Functional quasi-likelihood regression models with smooth random effects

November 2002

Jeng-Min Chiou

*Division of Biostatistics and Bioinformatics, National Health Research Institutes,
Taipei 115, Taiwan, R.O.C. Email: jmchiou@nhri.org.tw*

Hans-Georg Müller

*Department of Statistics, University of California, Davis, CA 95616, U.S.A.
Email: mueller@wald.ucdavis.edu*

Jane-Ling Wang

*Department of Statistics, University of California, Davis, CA 95616, U.S.A.
Email: wang@wald.ucdavis.edu*

Address for correspondence: Hans-Georg Müller, Department of Statistics, University of California at Davis, One Shields Avenue, Davis, CA 95616-8705, USA.

E-mail: mueller@wald.ucdavis.edu

Summary. We propose a class of semiparametric functional regression models to describe the influence of vector-valued covariates on a sample of response curves. Each observed curve is viewed as the realization of a random process, composed of an overall mean function and random components. The finite-dimensional covariates influence the random components of the eigenfunction expansion through single-index models that include unknown smooth link and variance functions. The parametric components of the single-index models are estimated via quasi-score estimating equations with link and variance functions being estimated nonparametrically. We obtain several basic asymptotic results. The proposed functional regression models are illustrated with the analysis of a data set consisting of egg-laying curves for 1000 female Mediterranean fruit flies (medflies).

Keywords: Estimating equations; Functional data analysis; Functional regression; Principal components; Semiparametric quasi-likelihood regression (SPQR); Single-index model; Smoothing.

1 Introduction

Methods for the analysis of data which consist of curves or similar infinite dimensional objects are in demand, as the collection of data of this type is becoming increasingly common. Curve data are being collected for instance in endocrinological (Brumback and Rice, 1998) and in growth studies (Gasser et al., 1984; Gasser and Kneip, 1995; Kneip and Gasser, 1992). Overviews of different methods and approaches that have been used for functions viewed as data can be found in the monographs by Ramsay and Silverman (1997, 2002).

The point of view we adopt here is that each observed curve is a realization of a stochastic process reflecting the random nature of the individual curves contained in a sample of curves. For curve data this approach was pioneered by Castro, Lawton and Sylvestre (1986) and was further developed in the seminal work of Rice and Silverman (1991), who emphasized the importance of smoothing. For many applications, it is of particular interest to model

the relation between covariates or associated variables as predictors and a response curve obtained for each subject. Incorporation of covariate effects in the form of time shifts was considered previously in Capra and Müller (1997), where a covariate Z is assumed to influence the time scale via a multiplicative scheme, $t' = \psi(Z)t$, for an unknown but smooth function ψ that can be estimated from the sample of curves. Other recent approaches for incorporating covariates into functional data models include the development of functional analysis of variance approaches (Brumback and Rice, 1998, Staniswalis and Lee, 1998) and adaptations of varying-coefficient regression models (Hoover et al., 1998). A related random effects model where the observed curves are assumed to be regression splines with random coefficients has been proposed in recent work of Rice and Wu (2001).

Our starting point is a sample of stochastic processes $X_i(t)$, $i = 1, \dots, n$, $t \in T$, where T is an interval. In practice, processes X_i are usually recorded on a grid of time points t_{ij} , $1 \leq j \leq m_i$, where we assume that these are equidistant or if not equidistant then on a regular grid generated by a design density. For each process X_i , we observe a vector of covariates or associated variables \mathbf{Z}_i , $\mathbf{Z}_i \in \mathbb{R}^p$, $p \geq 1$. The observations $\{\mathbf{Z}_i, X_i(\cdot)\}$ are assumed to be i.i.d.. The processes $X_i(\cdot)$ assume the role of response curves, which are influenced by predictors \mathbf{Z}_i . It is assumed that the overall mean function $\mu(\cdot)$ and the eigenfunctions or principal component functions $\rho_k(\cdot)$, $k = 1, 2, \dots$, defined via the autocovariance structure of processes X_i (see Appendix A for details) do not depend on the covariates \mathbf{Z}_i , whereas the principal components (also called principal component scores) and the conditional means of the processes X_i do. The case of a high-dimensional covariate vector \mathbf{Z}_i is handled by reducing the covariate vector to a single index. This single index is then embedded in a semiparametric quasi-likelihood regression (SPQR) model. We use the quasi-likelihood with unknown link and variance function estimation or QLUE approach (see Chiou and Müller, 1998), which provides an estimating equation with nonparametrically estimated components, asymptotic distribution theory and an algorithm for fitting the parametric and nonparametric components. We demonstrate the effectiveness of the proposed semiparametric approach

through fertility data where the sample of curves consists of the daily egg production observed individually for each of 1,000 female Mediterranean fruit flies (medflies); for an illustration, see Figure 1.

The paper is organized as follows: The proposed *functional smooth random effects model* for modeling covariate effects on curve data is presented in Section 2. The estimation of the components of the model, smoothing procedures and asymptotic results are discussed in Section 3. Applying this approach to the egg-laying data is the theme of Section 4, which also includes a discussion of various practical issues such as choice of the number of eigenfunctions and smoothing parameters. Additional technical material such as process representations in terms of eigenfunctions, proofs and details on the estimating equations for semiparametric quasi-likelihood regression (SPQR) can be found in appendices A-C.

2 Functional smooth random effects model

Principal components analysis of data vectors is a commonly used multivariate technique. Analysis of a sample of curves may be carried out in a similar way. Castro, Lawton and Sylvestre (1986) defined the concept of best K -dimensional linear model for stochastic processes and showed that this model corresponds to the first K terms of the Karhunen-Loève expansion, which is based on the eigenfunctions of the covariance kernel of the process. Rice and Silverman (1991) discussed the application of smoothing techniques for implementing this concept in curve data analysis, and proposed a leave-one-curve-out cross-validation technique to select smoothing parameters.

The predictor variable $\mathbf{Z} \in \mathbb{R}^p$ is observed along with the process $X(t), t \in T$, where T is an interval. We assume that \mathbf{Z} influences the random components A_k in the following class of *functional smooth random effects models*:

(M) Given a smooth function $\mu(t), t \in T$, random variables $A_k, k = 1, 2, \dots$, and eigenfunc-

tions $\rho_k : T \rightarrow \mathbb{R}$, let

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} A_k \rho_k(t),$$

and assume for the observed random curves, conditional on the covariates,

$$\mu_{\mathbf{Z}}(t) = E\{X(t)|\mathbf{Z}\} = \mu(t) + \sum_{k=1}^{\infty} E(A_k|\mathbf{Z})\rho_k(t).$$

The following assumptions are made:

(M1) There exists an integer $K < \infty$ such that the eigenvalues $\lambda_k = E(A_k^2)$ (see (17) in Appendix A) satisfy $\lambda_k < \infty$ for $k = 1, \dots, K$, and $\lambda_k = 0$, for $k > K$,

implying that only the first K terms of the expansion (M) matter.

(M2) The eigenfunctions $\rho_k(\cdot)$ are orthonormal in $L^2([0, T])$ and are twice continuously differentiable.

(M3) For all k, ℓ , using the Kronecker symbol $\delta_{k\ell} = 1$ if $k = \ell$, $\delta_{k\ell} = 0$ if $k \neq \ell$,

$$E(A_k) = 0 \quad \text{and} \quad \text{cov}(A_k, A_\ell) = \lambda_k \delta_{k\ell}.$$

From (M3), we find for the conditional covariance (and setting $s = t$, also conditional variance)

$$\text{cov}\{X(s), X(t)|\mathbf{Z}\} = \sum_{k,\ell} \text{cov}(A_k, A_\ell|\mathbf{Z}) \rho_k(s)\rho_\ell(t). \quad (1)$$

(M4) Each of the conditional link functions for the random effects is a smooth function of a single index formed from the data; i.e., there exists a parameter vector $\beta_k \in \mathbb{R}^p$, $\|\beta_k\| = 1$ and a twice continuously differentiable function $\alpha_k : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\alpha_k(\beta_k' \mathbf{z}) = E(A_k|\mathbf{Z} = \mathbf{z}),$$

for $k = 1, 2, \dots$

These single index models serve as a device for dimension reduction.

The proposed *functional smooth random effects model* is a natural extension of the classical Karhunen-Loève representation for stochastic processes (see (18) in Appendix A). It incorporates the influence of the covariate effects \mathbf{Z} through the conditional distribution of the principal components scores, which are the random effects in the model, via the smooth regression relation (M4).

The link functions α_k in (M4) are of intrinsic interest as they, jointly with the eigenfunctions ρ_k , determine the nature of the dependence of the process X on the predictors \mathbf{Z} . For a given eigenfunction, an interpretation in terms of the underlying subject-matter problem may be found by checking the processes which have large principal components for this eigenfunction; these “principal component scores” are in turn determined by the shape of the eigenfunction. Based on an interpretation of the k -th eigenfunction ρ_k , the link function α_k then reveals the changes in relative importance of this eigenfunction in the representation of (M4) and (M) as the covariate \mathbf{Z} varies.

We conclude this section with three remarks. First, the conditional covariance structure of the model as given in (1) is quite adaptable, while keeping the dimension of the nonparametric components low so as to avoid the “curse of dimensionality” associated with high-dimensional nonparametric smoothing.

Second, if the domain is $T = (t_0)$, a singleton, then X is just a real random variable and there is only one eigenfunction which is $\rho_1(t_0) = 1$. In this case, Model (M) collapses to a semiparametric regression model with an unknown link function $g(\cdot)$ and a parameter vector $\theta \in \mathbb{R}^p$ such that $E\{X(t_0)|\mathbf{Z} = \mathbf{z}\} = g(\theta'\mathbf{z})$, which is determined by the joint distribution of $(\mathbf{Z}, X(t_0))$. This connects our approach to quasi-likelihood regression models with unknown link and variance functions (see Appendix C). In this sense, model (M) provides a natural extension of *semiparametric quasi-likelihood regression* (SPQR) to the case where the responses are functions.

Third, consider a modified model with an explicit error process, given by $\tilde{X}(t) = X(t) +$

$\varepsilon(t)$, where the error process $\varepsilon(t)$ is independent of X and \mathbf{Z} and satisfies $E\{\varepsilon(t)\} = 0$, $cov\{\varepsilon(s), \varepsilon(t)\} = \sigma(s, t)$, where $\sigma(\cdot, \cdot)$ is square integrable. This implies that the processes \tilde{X} have eigenfunctions $\tilde{\rho}_k$, $k = 1, 2, \dots$ and that there exist uncorrelated zero mean random variables \tilde{A}_k such that $\sum \tilde{A}_k \tilde{\rho}_k(t) = \sum A_k \rho_k(t) + \varepsilon(t)$. Therefore, model (M) still applies, based on random effects \tilde{A}_k and eigenfunctions $\tilde{\rho}_k$, with corresponding altered interpretations.

3 Estimation of the model components

3.1 Preliminaries

For purposes of estimation and prediction, we have assumed that the number K of eigenfunctions with non-zero eigenvalues in model (M) is finite. If in reality there exist more than the assumed number of K components, the target is the projection on the space spanned by the first K eigenfunctions. We also assume from now on that the domain T of the processes is $T = [0, \tau]$, and that discretized measurements $X_i(t_{ij})$ of processes $X_i(\cdot)$ are available, $i = 1, \dots, n$ and $j = 1, \dots, m_i$. We develop our arguments for equi-spaced designs, for which $t_{ij} = t_j$, $j = 1, \dots, m$ and the t_j are generated by a smooth design density (see, e.g., Müller, 1984), but extensions to situations that are not equi-spaced and for which the number of measurements varies from subject to subject can be covered by adding a presmoothing step, if the number of measurements m_i for the i -th subject satisfies $0 < c_1 \leq m_i/m \leq c_2 < \infty$, $i = 1, \dots, n$ for constants $c_1, c_2 > 0$.

Since smoothing plays a central role in the proposed approach, we briefly discuss some relevant issues. Given any scatterplot data $(W_i, Y_i)_{i=1, \dots, s}$, with underlying regression function $m(w) = E(Y|W = w)$, we define the smoothed estimate $\hat{m}(\cdot)$ by $\hat{m}(w) = S\{w, b, (W_i, Y_i)_{i=1, \dots, s}\}$. Here, s is the number of data in the scatterplot and b is the smoothing parameter of the smoother S , which is evaluated at the argument w .

For our analysis, it does not matter which particular smoothing method is applied, as long as some basic consistency properties are satisfied and the smoother is linear in the data.

Kernel estimators, smoothing splines or local polynomial fitting by locally weighted least squares are among the possible choices. We choose here the locally weighted least squares smoother, denoted by S_L , fitting local lines to the data. A formal definition is

$$S_L\{w, b, (W_i, Y_i)_{i=1, \dots, s}\} = \arg \min_{a_0} \left\{ \min_{a_1} \sum_{i=1}^s K\left(\frac{w - W_i}{b}\right) [Y_i - \{a_0 + a_1(w - W_i)\}]^2 \right\}.$$

This means that the value of the smoother at the argument w is obtained as the estimated intercept of a regression line at w , where the line is fitted to those data falling into the window $[w - b, w + b]$ by weighted least squares. Here, $K(\cdot) \geq 0$ is a nonnegative kernel function, common choices being $K(x) = (1 - x^2) 1_{(|x| \leq 1)}$ or $K(x) = e^{-x^2/2}$.

3.2 Estimating overall means and eigenfunctions

A crude estimate of the overall mean function μ can be obtained by smoothing pointwise averages $\tilde{\mu}(\cdot)$. For equi-spaced designs, these pointwise averages are

$$\tilde{\mu}(t_j) = \frac{1}{n} \sum_{i=1}^n X_i(t_j), \quad (2)$$

given n sample curves X_1, X_2, \dots, X_n . To obtain a smooth function in t , we may add a smoothing step with a small bandwidth b_μ ,

$$\hat{\mu}(t) = S[t, b_\mu, \{t_j, \tilde{\mu}(t_j)\}_{j=1, \dots, m}]. \quad (3)$$

For the case of designs that are not equi-spaced, a pre-smoothing step can be carried out first with the aim of sampling all curves at the same time points.

Let Σ be the sample $m \times m$ variance-covariance matrix of the stochastic process X , with elements

$$v_{rs} = \frac{1}{n} \sum_{i=1}^n \{X_i(t_r) - \hat{\mu}(t_r)\} \{X_i(t_s) - \hat{\mu}(t_s)\}, \quad 1 \leq r, s \leq m. \quad (4)$$

Denote by $\tilde{\rho}_k = \{\tilde{\rho}_k(t_1), \tilde{\rho}_k(t_2), \dots, \tilde{\rho}_k(t_m)\}$ the eigenvector corresponding to the k -th largest eigenvalue of the matrix Σ . Following the method used in Capra and Müller (1997), which

is a variant of a proposal of Rice and Silverman (1991), a smooth estimate $\hat{\rho}_k(\cdot)$ of the eigenfunction $\rho_k(\cdot)$ is obtained simply by smoothing the vector $\tilde{\rho}_k$,

$$\hat{\rho}_k(t) = S[t, b_{\rho_k}, \{t_j, \tilde{\rho}_k(t_j)\}_{j=1, \dots, m}], \quad k = 1, \dots, K. \quad (5)$$

3.3 Estimating random effects

Estimation of the link functions α_k and parameter vectors β_k is based on the estimated overall mean function $\hat{\mu}$ and eigenfunctions $\hat{\rho}_k$. A discrete approximation of $A_k = \langle \rho_k, X - \mu \rangle$ (see (19)), then motivates the updated random effects estimates

$$\hat{A}_{ik} = \frac{\Delta}{m_i} \sum_{j=1}^{m_i} \{X_i(t_{ij}) - \hat{\mu}(t_{ij})\} \hat{\rho}_k(t_{ij}), \quad k = 1, \dots, K, \quad (6)$$

where $\Delta = t_{ij} - t_{i(j-1)}$. To update estimates $\hat{\alpha}_k(\hat{\beta}'_k \mathbf{Z})$, for $\mathbf{Z} \in \mathbb{R}^p$ and $p > 1$, we adopt

$$\{\hat{\beta}_k, \hat{\alpha}_k(\cdot)\} = QLU E\{(\mathbf{Z}_i, \hat{A}_{ik})_{i=1, \dots, n}\}, \quad (7)$$

using semiparametric quasi-likelihood regression (SPQR) in the form of the QLUE approach of Chiou and Müller (1998) which is summarized in Appendix C. For the special case where the covariate \mathbf{Z} is one-dimensional, we may obtain $\hat{\alpha}_k(\cdot)$ by a simple nonparametric regression step. In this case, there is no parameter vector, and we can regress the \hat{A}_{ik} nonparametrically on the Z_i , $i = 1, \dots, n$, to obtain the smooth function estimates

$$\hat{\alpha}_k(v) = S\{v, b_{\alpha_k}, (Z_i, \hat{A}_{ik})_{i=1, \dots, n}\}. \quad (8)$$

3.4 Prediction and estimation algorithm

For prediction, given a vector of covariates \mathbf{Z} as described in Section 2, we substitute the unknown functions μ , $\{\alpha_k\}_{k=1, \dots, K}$ and $\{\rho_k\}_{k=1, \dots, K}$, in the model with the estimates $\hat{\mu}$, $\hat{\alpha}_k$ and $\hat{\rho}_k$, respectively. The prediction for $\mu_Z(t) = E\{X(t)|\mathbf{Z}\}$ is then

$$\hat{\mu}_Z(t) = \hat{\mu}(t) + \sum_{k=1}^K \hat{\alpha}_k(\hat{\beta}'_k \mathbf{Z}) \hat{\rho}_k(t). \quad (9)$$

The estimation procedure of the model components is summarized as follows:

- (E1) Obtain the estimates $\hat{\mu}(t)$ of the overall mean function $\mu(t)$ by (2) and (3).
- (E2) Obtain the estimates $\hat{\rho}_k(t)$ of the eigenfunctions $\rho_k(t)$ by (5).
- (E3) Obtain the link function fits $\hat{\alpha}_k(\cdot)$ and the parameter vector estimates $\hat{\beta}_k$ via (7), the QLUE approach.

The smoothing and QLUE steps require the choice of a smoothing parameter. The choice used in the QLUE algorithm is based on special features of quasi-likelihood, see Chiou and Müller (1998). Smoothing parameter selection for the function estimates $\hat{\rho}_k$ and $\hat{\mu}_0$ requires other methods. We will discuss practical choices coupled to the leave-one-curve-out technique in Sections 4.2 and 4.4.

3.5 Some basic asymptotic results

Following arguments of Pezzulli and Silverman (1993), it is possible to obtain consistency properties for eigenfunction estimates $(\hat{\rho}_k)_{k=1,\dots,K}$, under mild regularity conditions. For a function $\varphi \in L^2(d\nu)$ on domain T , set $\|\varphi\| = \{\int_T \varphi^2(t)d\nu(t)\}^{1/2}$. We consider here a simplified version of model (M) where $\mathbf{Z} \in \mathbb{R}^1$ and it follows that $\beta_k = 1$ for $k = 1, \dots, K$.

For the smoother S used in the smoothing steps, we require the following two minimal properties, which are satisfied by virtually all commonly used smoothing methods.

- (S1) Given i.i.d. random pairs $(W_i, Y_i)_{i=1,\dots,s}$, from a distribution with regression function $m(w) = E(Y|W = w)$, if the regression function $m(\cdot)$ is twice continuously differentiable, and the p.d.f. f_W of the W_i 's is continuous at a point w and satisfies $f_W(w) > 0$, then

$$S\{w, b, (W_i, Y_i)_{i=1,\dots,s}\} - m(w) = O_p(\tau_n),$$

for a sequence $\tau_n \rightarrow 0$, which depends on the particular smoother and smoothing parameter choice.

(S2) The smoother is linear in the data Y_i ,

$$S\{w, b, (W_i, Y_i)_{i=1, \dots, s}\} = \sum_{i=1}^s G_i(w) Y_i$$

for (possibly random) weight functions G_i , independent of Y_i , and satisfies

$$\left\{ \sum_{i=1}^s G_i^2(w) \right\} \left\{ \sum_{i=1}^s 1_{\{G_i(w) \neq 0\}} \right\} = O_p(1).$$

Examples of smoothers satisfying (S1), (S2) are local linear fits or kernel smoothers, among others.

Theorem 1. If the smoothers S employed in estimating the nonparametric model components satisfy (S1) and (S2), and the covariate Z has a density which is continuous and positive at any given point z , then

$$\|\hat{\mu} - EX\| = O_p(n^{-1/2}). \quad (10)$$

If in addition $\|\hat{\rho}_k - \rho_k\| = O_p(\sigma_n)$, $k = 1, \dots, K$, for a sequence $\sigma_n \rightarrow 0$, then

$$|\hat{\alpha}(z) - \alpha(z)| = O_p(\sigma_n + \tau_n + n^{-1/2} + m^{-1}). \quad (11)$$

The proof of this result is in Appendix B. Consistency requires, not surprisingly, that both the number n of observed sample curves as well as the number of measurements m taken per curve satisfy $n \rightarrow \infty$, $m \rightarrow \infty$.

We note that a uniform result is possible for (11),

$$\sup_z |\hat{\alpha}(z) - \alpha(z)| = O_p\{\nu_n(\sigma_n + \tau_n + n^{-1/2} + m^{-1})\},$$

if instead of (S2) we require

$$\sup_z \left\{ \sum_{i=1}^n G_i^2(z) \right\} \left\{ \sum_{i=1}^n 1_{\{G_i(z) \neq 0\}} \right\} = O_p(\nu_n).$$

A typical value which is achieved by most smoothing methods is $\nu_n = (\log n/nb)^{-1/2}$, where b is the bandwidth-equivalent. Uniform consistency therefore requires in addition that $\nu_n(\sigma_n +$

$\tau_n + n^{-1/2} + m^{-1}) \rightarrow 0$. Under certain assumptions, Pezzulli and Silverman (1993) have shown that $\|\hat{\rho}_k - \rho_k\| = o_p(1)$, which leads to the consistency result $|\hat{\alpha}(z) - \alpha(z)| = o_p(1)$.

These results can be extended to the case of multivariate predictors, using the asymptotic consistency properties of QLUE as described in Chiou and Müller (1998). For the following result on the QLUE estimators for the unknown link functions α_k and parameter vectors β_k we assume for simplicity that the true μ and ρ_k are known, so that the $\{A_{ik}\}$ play the role as the usual “response” variables in the semiparametric quasi-likelihood (SPQR) step.

Theorem 2. Suppose that the overall mean function μ and the eigenfunctions ρ_k in model (M) are given. Then, under regularity conditions, the following event holds for each $k, k = 1, \dots, K$ with probability $1 - \delta$, for a given arbitrarily small $\delta > 0$: There exist QLUE estimates $\hat{\beta}_k$ of parameter vectors β_k , satisfying $\|\hat{\beta}_k\|=1$, and $\hat{\alpha}_k$ of link functions α_k , such that, as $n \rightarrow \infty$,

$$\sup_z |\hat{\alpha}(z) - \alpha(z)| = o_p(1) , \quad (12)$$

$$\sqrt{n} (f(\hat{\beta}_k) - f(\beta_k)) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, [(Df)(\beta_k)] \Sigma^{-1} [(Df)(\beta_k)]^\top) , \quad (13)$$

where $\Sigma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n D_i^\top V_i^{-1} D_i$ with $D_i = (\partial \alpha(\beta_k^\top \mathbf{x}_i) / \partial \beta_k^\top)$, $f(\beta_k) = (f_1(\beta_k), \dots, f_{p-1}(\beta_k))^\top$ with $f_j(\beta_k) = \beta_{kj} / \|\beta_k\|$, and $(Df)(\beta_k) = \left(\frac{\partial f_i(\beta_k)}{\partial \beta_{kj}} \right)_{1 \leq i \leq p-1, 1 \leq j \leq p}$, defined as a $(p-1) \times p$ matrix.

We note that the degrees of freedom for the regression parameter estimates as well as rank of the corresponding asymptotic covariance estimates are reduced by one due to the identifiability constraint in (M4). Details of the proof of the theorem including the adjustment of the asymptotic covariance estimates are similar to Chiou and Müller (1998, 2002) and are therefore omitted.

4 Application to egg-laying curves for Medflies

4.1 Background

Recently, increased interest in research on aging and quantitative biodemography has focused on the relationship between reproductive and aging patterns. Evolution is driven by reproductive success, and the connection between the evolution of aging and reproduction is intriguing not least because it has proved to be elusive. Unraveling this connection may aid our understanding of the aging process. Patterns of reproduction are typically inferred from experiments which involve large experimental cohorts. The study of reproductive behavior of large cohorts of medflies (*Ceratitits capitata*) has led to several interesting findings. For instance, an interaction of nutritional and gender effects results in a reversal of the female-male life expectancy differential under protein deprivation, as reported in Müller et al. (1997), and two different modes of aging and reproduction in conjunction with diet were identified in Carey et al. (1998b). It has been conjectured that reproductive activity has a negative effect on longevity, and the term “cost of reproduction” was coined by Partridge and Harvey (1985). Flies are ideal for the study of reproduction and longevity, as large cohorts can be reared with recordings of both daily egg-laying and survival. We focus here on the relationship between patterns in the reproduction curves which are viewed as dependent data, with the total number of eggs produced by a fly and its lifetime as predictors. Total number of eggs produced is a measure of reproductive success, and it is of interest to find out how reproductive patterns are associated with overall reproductive success and lifetime.

The experiment which provides the data for our analysis was carried out in 1992-1995 at the medfly mass rearing and sterilization facility (Moscamed) at Metapa, Chiapas, Mexico and consisted of $n = 1,000$ female medflies for which daily egg reproduction was recorded. More details about the biological features of the experiment can be found in Carey et al. (1998a).

Let $X_i(\cdot)$ denote the egg-laying curve of the i th fly and \mathbf{Z}_i the bivector of lifetime and

total number of eggs. Randomly subsampled egg-laying curves are visualized in Figure 1, displaying 20 curves for each quartile of total number of eggs. From the initial sample of 1,000 medflies, those who did not lay any eggs were discarded as those curves have constant value zero, leaving egg-laying data for $n = 936$ medflies in the analysis. We restricted the analysis to the first 50 days of egg-laying, as we found large variability of egg-laying at higher ages which would entirely dominate the eigenfunctions, and as an earlier analysis in Carey et al. (1998a) had shown that total number of eggs as a function of lifetime showed a marked change-point at 51 days. Thus, the egg-laying curves $X_i(t)$, $i = 1, \dots, 936$, are considered as realizations of a stochastic process on $T = [0, \tau]$ with $\tau = 50$ days.

As many of the flies died before 50 days, it is of interest to consider whether informative dropout occurs in the sense that the mean function of egg-laying is affected by the pattern of dropouts. An approximate parametric model of the egg-laying process was developed and used in Müller et al. (2001) to define a putative egg-laying random process after time of death, and in particular a remaining egg-laying potential random function. While interesting associations were established between lifetime and remaining egg-laying potential at the time of death, the association between lifetime and the imputed remaining egg-laying process after death was found to be at best very weak, providing some justification for the assumption that the dropout is non-informative in this example.

4.2 *Eigenfunctions*

The first four estimated eigenfunctions $\hat{\rho}_k$ (5) are shown in Figure 2 with the proportion of variance explained for each principal component indicated in the legend. The corresponding eigenvalues range from 0 to above 6,000 which indicates large variation, in accordance with the variability seen in Figure 1. The first eigenfunction is roughly similar to the overall mean shape of egg-laying curves, with a steep rise in the first 10 days and a subsequent relatively flat downwards pointing slope.

If we suppose that in fact $\rho_1(t) = \mu(t)$, and that only one principal component matters,

we obtain the multiplicative effects model

$$E\{X(t)|\mathbf{Z} = \mathbf{z}\} = \mu(t)\{1 + E(A_1|\mathbf{Z} = \mathbf{z})\}$$

and

$$var\{X(t)|\mathbf{Z} = \mathbf{z}\} = \mu^2(t)var\{A_1|\mathbf{Z} = \mathbf{z}\}.$$

Conditionally on $\mathbf{Z} = \mathbf{z}$, we therefore observe quasi-Gamma type behavior. Unconditionally,

$$E\{X(t)\} = \mu(t), \quad var\{X(t)\} = \mu^2(t)\lambda_1,$$

and the “first mode of variation” is “in the direction of” the mean function $\mu(\cdot)$. Thus, unconditionally, we obtain a constant coefficient of variation model in t , $\sqrt{\lambda_1}$ being the coefficient of variation.

We find that the second eigenfunction differentiates this shape into a very early peak at around 5 days and a subsequent much faster decline, leaving room for further peaks later on. The third eigenfunction depicts a broad peak at later ages, and the higher order eigenfunctions resolve this into a series of increasingly complex oscillations.

The mean function $\hat{\mu}$ (3) does not depend on the covariate and is shown in Figure 3. The smoothing parameter is determined by minimizing the squared prediction error obtained by the leave-one-curve-out technique,

$$PE = \sum_{i=1}^n \sum_{j=1}^{m_i} \{\hat{X}_i^{(-i)}(t_{ij}) - X_i(t_{ij})\}^2 / nm_i, \quad (14)$$

where $\hat{X}_i^{(-i)}$ is the leave-one-curve-out function estimate obtained by omitting the i th sample curve in the estimation. Since for the classical Karhunen-Loève model without covariates $\hat{X}_i^{(-i)}(t) = \hat{\mu}^{(-i)}(t) = \hat{\mu}^{(-i)}(t, b)$, we can compute the optimal prediction smoothing parameter $\hat{b}_\mu = \arg \min_b PE(b)$; this produced $\hat{b}_\mu = 2.5$ days.

4.3 Total number of eggs as a single covariate

We first consider the case of a single covariate. The covariate Z is the total number of eggs, which is the predictor for the random components A_k . In addition to the estimation

of the eigenfunctions ρ_k and the mean function μ we need to obtain estimates $\hat{\alpha}_k$ (8) for the random effect link functions α_k . Implementing these steps leads to the estimated functions $\hat{\alpha}_k$, shown in Figure 4 for $k = 1 - 4$. We find that $\hat{\alpha}_1$ is monotone increasing throughout and almost linear, while $\hat{\alpha}_2$ is slightly increasing at the beginning and then monotone decreasing throughout with increasingly negative slope. The functions $\hat{\alpha}_k$ of higher order ($k = 3, 4$ and higher) appear flat, indicating that the covariate only affects the scores for the first two eigenfunctions. Indeed our final model will not use all four random components.

Given function estimates $\hat{\mu}(\cdot)$ and $\{\hat{\alpha}_k(\cdot)\}_{k=1,\dots,K}$, we obtain the fitted surface

$$\hat{\mu}_Z(t) = \hat{\mu}(t) + \sum_{k=1}^K \hat{\alpha}_k(Z) \hat{\rho}_k(t), \quad (15)$$

according to (9). The fitted surface $\hat{\mu}_Z(t)$ (15), shown in the upper panel of Figure 5, exhibits an interesting sloping ridge pattern, and the cross-sections, in the lower panel of Figure 5, reveal that while the peaks are slightly drifting to the right as the total number of eggs increases, the graphs are, for the most part, being shifted upwards as the total number of eggs increases. We note that the number of eigenfunctions is chosen as $K = 2$ for Figure 5. This choice is based on visual inspection of the functions $\hat{\alpha}_k$ in Figure 4, as well as the predictive quality of the fitted model.

The leave-one-curve-out prediction error (14) is a useful quantification of the predictive quality of a model. The leave-one-curve-out predictors are

$$\hat{\mu}_{Z_i}^{(-i)}(t) = \hat{\mu}^{(-i)}(t) + \sum_{k=1}^K \hat{\alpha}_k^{(-i)}(Z_i) \hat{\rho}_k^{(-i)}(t). \quad (16)$$

Here, $\hat{\mu}^{(-i)}$, $\hat{\rho}_k^{(-i)}$ and $\hat{\alpha}_k^{(-i)}$ are all obtained in the same way as before, but omitting the observed process X_i itself. This functional form of the cross-validation sum of squares was introduced by Rice and Silverman (1991). Prediction errors in dependency on the number of fitted eigenfunctions are shown in Table 1, supporting the choice $K = 2$. The leave-one-curve-out prediction error for a model without the covariate effect, that is the model specified by (18) in Appendix A, was found to be 462.54. The reductions in prediction error when

Table 1: Prediction errors (PE) (14) of fitted Model (16) with K eigenfunctions.

| K | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|------|--------|--------|--------|--------|--------|--------|--------|--------|
| PE | 333.71 | 313.22 | 315.22 | 316.41 | 315.64 | 315.62 | 315.62 | 315.49 |

including covariates demonstrate the effectiveness of including the covariate effect in Model (M).

4.4 *The case of multiple covariates*

For the case of multiple covariates, we consider both total number of eggs laid and lifetime as the predictors for the random components A_k . The estimation procedure described in Section 3 was implemented for standardized covariates. The dimension for the random components used in the analysis is again chosen as $K = 2$. The first four estimated random effects link functions $\hat{\alpha}_k$ are in Figure 6.

The estimated coefficients $\hat{\beta}_1 = (\hat{\beta}_{11}, \hat{\beta}_{12})$ and $\hat{\beta}_2 = (\hat{\beta}_{21}, \hat{\beta}_{22})$ for the linear predictors in functions α_1 and α_2 are presented in Table 2. These estimates are obtained from the semiparametric quasi-likelihood regression approach implemented with QLUE. The associated standard errors are useful for asymptotic inference. For the linear predictor β_1 , total number of eggs is relatively more important than lifetime, while both predictors are significant and of opposite sign, forming a contrast in the covariates. For linear predictor β_2 , the influence of the two predictors is about the same, forming an average of the covariates, and both predictors are significant.

The effects of individual predictors on the response are a consequence of the coefficients listed in Table 2, and the eigenfunctions in Figure 2. For the particular application at hand we propose here a criterion for the necessary smoothing parameter selection in for the estimation of the link functions that is easy to implement and has good practical properties. It is based

Table 2: QLUE for the random components $\alpha_k(\beta'_k \mathbf{V})$ corresponding to Figures 6.

| Function | Coefficient | QLUE (Std.Err.) | Bandwidth | G.O.F. |
|------------|-------------------------|-----------------|--------------|-------------|
| α_1 | β_{11} (TotalEgg) | 0.8842 (0.0031) | 83.78 (Var.) | D = 940.36 |
| | β_{12} (Lifetime) | -.4671 (0.0059) | 1.16 (Link) | P = 916.20 |
| α_2 | β_{21} (TotalEgg) | 0.6730 (0.0520) | 28.22 (Var.) | D = 1250.75 |
| | β_{22} (Lifetime) | 0.7397 (0.0471) | 1.91 (Link) | P = 936.60 |

on pseudo-likelihood (PL; see Davidian and Carroll, 1988),

$$PL\{y; \mu(\beta), \sigma^2(\mu)\} = -\frac{1}{2} \sum_{i=1}^n \log \{2\pi\sigma^2(\mu_i)\} - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\sigma^2(\mu_i)},$$

where y , μ and σ^2 , respectively, represent the vectors of response, mean and variance. Here μ and σ^2 are replaced by their estimates $\hat{\mu}$ and $\hat{\sigma}^2$ in the pseudo-likelihood and optimal bandwidths are chosen by maximizing the pseudo-likelihood $PL\{y; \mu, \sigma^2\}$.

The resulting bandwidths for our data analysis are shown in the fourth column of Table 2. In addition, the column labeled G.O.F. in Table 2 presents the goodness-of-fit statistics for semiparametric quasi-likelihood regression. Here, we provide “nonparametric” quasi-deviance (D) and Pearson statistic (P), which are asymptotically chi-square distributed with the degrees of freedom approximately equal to the number of observations minus the number of parameters to be estimated in the linear predictor. The values in Table 2 do not indicate lack of fit, as $D \approx P$ (compare Chiou and Müller, 1998, for more details).

5 Discussion and concluding remarks

We have developed an approach to extend principal component analysis for curve data to the situation when covariates are present. This approach extends previous work of Rice and Silverman (1991) and yields a flexible functional regression model, the *functional smooth random effects model*. This model allows for substantial flexibility in mean regression and

heteroscedasticity structure. Flexibility is retained for high dimensions through the single index feature of the model. Through an example of egg-laying curves for 1,000 female medflies, this method is shown to be effective in reducing the prediction errors. The proposed approach is simple both conceptually and computationally. The nonparametric components of the semiparametric quasi-likelihood regression or SPQR model are structured in such a way that only one-dimensional smoothing steps are needed.

Our main results and techniques have been derived for the case of equi-spaced designs, where the available measurements for the curve data are made on a regular grid. Non-equispaced data are clearly important in practice. If a design is "mildly" non-equidistant in the sense that there are no persistent gaps, the methods as described in Section 3.1 can still be applied; for the case of persistent gaps in measurements and sparse designs, extensions of the proposed method need to be developed.

An extension to the practically relevant case where the predictor is a time-varying function rather than a time-invariant vector is possible through various approaches. One such approach is to discretize a covariate function $Z(t)$ into a (possibly high-dimensional) vector $Z(s_1), \dots, Z(s_M)$ by specifying an equidistant grid of M points s_1, \dots, s_M and then fitting the functional smooth random effects model as described above, using these vectors as covariates. An alternative approach is to project the random covariate functions onto their first K' eigenfunctions for some suitable integer K' and to use the resulting principal component score vectors as K' -dimensional covariate vectors in the functional smooth random effects model.

In the proposed approach, for the case of a one-dimensional covariate, acting only on the random effect, all model components can be estimated easily and consistently in one step, and we provide an asymptotic consistency result. The extension to the case of multivariate predictors is achieved by combining these ideas with a single-index quasi-likelihood approach for multivariate and high-dimensional regression, using the QLU technique. This allows to include both unknown link and variance functions. Our implementation also provides

for automatic choices of the smoothing parameters, so that only the covariates need to be identified and no specifications need to be made regarding the nature of the distribution of the random effects or the shape of the link functions.

Acknowledgements

We wish to thank Professor J. Carey for making the medfly fecundity data available to us. We are also grateful for most helpful remarks provided by two referees, an Associate Editor and the Joint Editor. This research was supported in part by National Health Research Institutes Grant BS-090-PP-07 , National Science Council Grant 89-2118-M-194-001, National Science Foundation Grants DMS-99-71602 and DMS-02-04869 and National Institutes of Health Grant 99-SC-NIH-1028.

Appendix A: Stochastic processes in L^2

Suppose that the observed curves X_1, X_2, \dots, X_n are i.i.d. realizations of a stochastic process X on a domain $T \subset \mathbb{R}$. The process is assumed to have mean and covariance functions

$$E\{X(t)\} = \mu(t), \text{ cov}\{X(s), X(t)\} = \gamma(s, t).$$

Given functions f and g , let $\langle f, g \rangle$ denote the $L^2(d\nu)$ inner product,

$$\langle f, g \rangle = \int_T f(t)g(t)d\nu(t),$$

where $d\nu$ is a measure on T which normally is chosen as the Lebesgue measure.

It is assumed that there exists an expansion of γ into orthonormal eigenfunctions $\rho_k(\cdot)$ on $L^2(d\nu)$ such that

$$\gamma(s, t) = \sum_{k=1}^{\infty} \lambda_k \rho_k(s) \rho_k(t), \tag{17}$$

with ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$. The eigenvalues are nonnegative due to the fact that the covariance kernel $\gamma(s, t)$ is symmetric and non-negative definite. The

orthonormal eigenfunctions satisfy $\langle \rho_i, \rho_i \rangle = 1$ and $\langle \rho_i, \rho_j \rangle = 0$ for $i \neq j$, and $\langle \gamma(s, \cdot), \rho_k \rangle = \lambda_k \rho_k(s)$, $s \in T$, $k = 1, 2, \dots$. The eigenvalues may then be expressed as $\lambda_k = \int_T \int_T \rho_k(s) \gamma(s, t) \rho_k(t) d\nu(s) d\nu(t)$, $k = 1, 2, \dots$. Under assumption (17), the stochastic process model then provides the following Karhunen-Loève representation for a randomly selected curve

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} A_k \rho_k(t), \quad (18)$$

where ρ_k is the k th eigenfunction. By Mercer's theorem (see Riesz and Nagy, 1990), the right hand side of (18) converges uniformly in $t \in T$. The random variables A_k correspond to the principal components, and are given by

$$A_k = \langle \rho_k, X - \mu \rangle. \quad (19)$$

The principal components A_k are uncorrelated random variables and satisfy

$$E(A_k) = 0, \quad \text{var}(A_k) = \lambda_k, \quad \sum_{k=1}^{\infty} \lambda_k < \infty,$$

i.e., the k -th eigenvalue corresponds to the variance of the k -th principal component. From (18), we find that under (17), $X(t)$ is composed of a mean function μ plus additive noise, the stochastic part.

The principal components A_k and basis functions ρ_k can be interpreted as defining the variation of the stochastic process about its mean function. According to (19), the random variable A_1 is the length of the projection of $X - \mu$ onto ρ_1 for each sample curve X , and $A_1 \rho_1$ explains the maximum amount of variation in X among all functions which involve a single real-valued random variable. Similarly, the function $A_2 \rho_2$ explains the maximum additional amount of process variation which is unexplained by $A_1 \rho_1$, and so forth for $k = 3, 4, \dots$.

Appendix B: Proof of Theorem 1

For (10), we observe

$$E(\|\tilde{\mu} - EX\|^2) = \frac{1}{n^2} \sum_{i,j,k,\ell} E(A_{ik}A_{j\ell}) \langle \rho_k, \rho_\ell \rangle = \frac{1}{n^2} \sum_{i,k} E(A_{ik}^2) \langle \rho_k, \rho_k \rangle = O(n^{-1}).$$

This implies $E(\|\tilde{\mu} - \mu\|) = O(n^{-1/2})$ by Jensen's inequality, and therefore $\|\tilde{\mu} - \mu\| = O_p(n^{-1/2})$. The assertion follows from (S2) and

$$\|\hat{\mu} - \mu\| \leq \|\tilde{\mu} - \mu\| \sup_x \sum_{i=1}^n G_i^2(x) \cdot \sum_{i=1}^n 1_{\{G_i(x) \neq 0\}}. \quad (20)$$

For (11), let $A_{ik}^{(1)} = \langle X_i - \hat{\mu}, \hat{\rho}_k \rangle$; $A_{ik}^{(2)} = \langle X_i - \mu, \rho_k \rangle$; $\hat{\alpha}_k^{(1)}(z) = S\{z, b_{\alpha_k}, (Z_i, A_{ik}^{(1)})_{i=1, \dots, n}\}$; $\hat{\alpha}_k^{(2)}(z) = S\{z, b_{\alpha_k}, (Z_i, A_{ik}^{(2)})_{i=1, \dots, n}\}$. Noting that $E(A_{ik}^{(2)}|Z) = \alpha_k(Z)$, the consistency assumption (S1) on the smoother implies $|\hat{\alpha}_k^{(2)}(z) - \alpha_k(z)| = O_p(\tau_n)$. Furthermore, observing the linearity assumption (S2),

$$\begin{aligned} |\hat{\alpha}_k^{(2)}(z) - \hat{\alpha}_k^{(1)}(z)| &= \left| \sum_{i=1}^n G_i(z) \{A_{ik}^{(2)} - A_{ik}^{(1)}\} \right| \\ &\leq \left\{ \sum_{i=1}^n G_i^2(z) \right\}^{1/2} \left\{ \sum_{i=1}^n (A_{ik}^{(2)} - A_{ik}^{(1)})^2 1_{\{G_i(z) \neq 0\}} \right\}^{1/2}. \end{aligned} \quad (21)$$

Note that

$$\begin{aligned} &\sup_i |A_{ik}^{(2)} - A_{ik}^{(1)}| \\ &\leq \sup_i |\langle X_i, \rho_k - \hat{\rho}_k \rangle| + |\langle \hat{\mu}, \hat{\rho}_k - \rho_k \rangle| + |\langle \rho_k, \hat{\mu} - \mu \rangle| \\ &= O_p(\|\hat{\mu} - \mu\| + \|\hat{\rho}_k - \rho_k\|), \end{aligned}$$

since K is finite, so that $\sup_i \|X_i\| < \infty$. Then (21) implies that $|\hat{\alpha}_k^{(2)}(z) - \hat{\alpha}_k^{(1)}(z)| = O_p(\|\hat{\mu} - \mu\| + \|\hat{\rho}_k - \rho_k\|)$, using (S2) in the same way as in (20). A second application of the same argument, using $\sup_i |A_{ik}^{(1)} - \hat{A}_{ik}| = O_p(m^{-1})$, which follows from the smoothness properties of $X_i, \hat{\rho}_k$ and $\hat{\mu}$, and (6) then implies the result.

Appendix C: Semi-parametric quasi-likelihood regression (SPQR)

Semiparametric quasi-likelihood regression (SPQR) is an estimation procedure for estimating the unknown model components: These consist of two smooth functions, the link function

$g(\cdot)$ and the variance function $\sigma^2(\cdot)$. and a vector of regression parameter β_0 , with $\|\beta_0\| = 1$. Given the observations y_i and the predictors \mathbf{x}_i , the model assumptions are

$$E(y_i) = g(\mathbf{x}_i^T \beta_0), \quad i = 1, \dots, n,$$

$$\text{var}(y_i) = \text{var}(\varepsilon_i) = \sigma^2\{g(\mathbf{x}_i^T \beta_0)\} = \sigma^2\{E(y_i)\}.$$

A three-stage iterative QLUE (quasi-likelihood with unknown link and variance function estimation) procedure was proposed by alternating the parametric and nonparametric estimation steps. The procedure can be summarized as follows, see Chiou and Müller (1998, 1999) for additional details.

Let $S^{(\nu)}\{z, h; (z_i, y_i)_{i=1, \dots, n}\}$, $\nu \geq 0$, be a generic notation for a nonparametric estimator for the ν th derivative of a regression function, $\frac{d^\nu}{dz^\nu} E(Y|Z = z)$, based on scatterplot data $(z_i, y_i)_{i=1, \dots, n}$. Here the z_i 's are design points, y_i 's are raw data to be smoothed, h denotes a bandwidth, and z is a target level at which the function is to be evaluated. Let g_ν be the ν th derivative of the link function g . The various updating steps are as follows.

1. *Nonparametric estimation step for link function.* For given $\hat{\beta}$, $\|\hat{\beta}\| = 1$, estimates of the link function and its first derivative are updated by

$$\hat{g}_\nu(t; \hat{\beta}) = S^{(\nu)}\{t, b_\nu; (\mathbf{x}_i^T \hat{\beta}, y_i)_{i=1, \dots, n}\}, \quad \nu = 0, 1.$$

2. *Nonparametric estimation step for variance function.* For given $\hat{\beta}$, $\|\hat{\beta}\| = 1$, and $\hat{g}_0(\cdot; \hat{\beta})$, an updated nonparametric variance estimate $\hat{\sigma}^2(\cdot)$ is obtained by

$$\hat{\sigma}^2(u) = S^{(0)}\{u, b; (\hat{\mu}_i, \hat{\varepsilon}_i^2)_{i=1, \dots, n}\}$$

where $\hat{\mu}_i = \hat{g}_0(\mathbf{x}_i^T \hat{\beta}; \hat{\beta})$ and $\hat{\varepsilon}_i^2 = (y_i - \hat{\mu}_i)^2$ are squared residuals which serve as the “raw” variance estimates and are based on a current model fit.

3. *Parametric estimation step.* For given $\hat{g}_0(\cdot; \hat{\beta})$, $\hat{g}_1(\cdot; \hat{\beta})$ and $\hat{\sigma}^2(\cdot)$, $\hat{\beta}$ is updated by solving the quasi-likelihood estimating equation with respect to β , inserting the current

estimates for link and variance function. The estimated estimating equation for the score is

$$U^*(\beta) = \sum_{i=1}^n \frac{y_i - \hat{g}_0(\eta_i; \beta)}{\hat{\sigma}^2\{\hat{g}_0(\eta_i; \beta)\}} \hat{g}_1(\eta_i; \beta) \mathbf{x}_i,$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$, $\eta_i = \mathbf{x}_i^T \beta$ and $\|\beta\| = 1$.

The iteration is run until convergence. Sliced inverse regression (SIR) (Li, 1991) estimates provide satisfactory starting values for β . To simplify the notation, we denote the estimated vector of regression parameters and the estimated link function obtained at convergence of the iteration by

$$\{\hat{\beta}, \hat{g}(\cdot)\} = QLU\{(\mathbf{x}_i, y_i)_{i=1, \dots, n}\}.$$

References

- Brumback, B. and Rice, J. (1998) Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *J. Am. Statist. Ass.*, **93**, 961-994.
- Capra, W. B. and Müller, H. G. (1997) An accelerated time model for response curves. *J. Am. Statist. Ass.*, **92**, 72-83.
- Carey, J., Liedo, P., Müller, H. G., Wang, J. L. and Chiou, J. M. (1998a) Relationship of age patterns of fecundity to mortality, longevity and lifetime reproduction in a large cohort of Mediterranean fruit fly females. *J. Gerontology*, **53A**, B245-B251.
- Carey, J., Liedo, P., Müller, H. G., Wang, J. L. and Vaupel, J. (1998b) Dual modes of aging in Mediterranean fruit fly females. *Science*, **281**, 996-998.
- Castro, P. E., Lawton, W. H., and Sylvestre, E. A. (1986) Principal modes of variation for processes with continuous sample curves. *Technometrics*, **28**, 329-337.
- Chiou, J. M. and Müller, H. G. (1998) Quasi-likelihood regression with unknown link and variance functions. *J. Am. Statist. Ass.*, **93**, 1376-1387.
- (1999) Nonparametric Quasi-likelihood. *Ann. Statist.*, **27**, 36-64.

- (2002) Quasi-likelihood regression with multiple indices and smooth link and variance functions. *Manuscript*.
- Davidian, M. and Carroll, R. J. (1988) A note on extended quasi-likelihood. *J. R. Statist. Soc. B*, **50**, 74-82.
- Gasser, T. and Kneip, A. (1995) Searching for structure in curve samples. *J. Am. Statist. Ass.*, **90**, 1179-1188.
- Gasser, T., Müller, H. G., Köhler, W., Molinari, L. and Prader, A. (1984) Nonparametric regression analysis of growth curves. *Ann. Statist.*, **12**, 210-229.
- Hart, J. P. and Wehrly, T. E. (1986) Kernel regression estimation using repeated measurements data. *J. Am. Statist. Ass.*, **81**, 1080-1088.
- Hoover, D., Rice, J., Wu, C. and Yang, L.-P. (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 809-822.
- Kneip, A. and Gasser, T. (1992) Statistical tools to analyze data representing a sample of curves. *Ann. Statist.*, **20**, 1266-1305.
- Li, K. C. (1991) Sliced inverse regression for dimension reduction. *J. Am. Statist. Ass.* **86**, 316-327.
- Müller, H. G. (1984) Optimal designs for nonparametric kernel regression. *Statist. Probab. Lett.*, **2**, 285-290.
- Müller, H. G., Carey, J.R., Wu, D., Liedo, P., Vaupel, J.W. (2001). Reproductive potential predicts longevity of female Mediterranean fruit flies. *Proceedings of the Royal Society B* **268**, 445-450.
- Müller, H. G., Wang, J. L., Capra, W. B., Liedo, P. and Carey, J. R. (1997) Early mortality surge in protein-deprived females causes reversal of sex differential of life expectancy in Mediterranean fruit flies. *Proc. Acad. Sci. USA*, **94**, 2762-2765.
- Partridge, L. and Harvey, P. H. (1985) Costs of reproduction. *Nature*, **316**, 20-21.
- Pezzulli, S. and Silverman, B. W. (1993) Some properties of smoothed principal components analysis for functional data. *Comput. Statist.*, **8**, 1-16.

- Ramsay, J. O. and Silverman, B. W. (1997) *Functional Data Analysis*. New York: Springer.
- Rice, J. A. and Silverman, B. W. (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc. B*, **53**, 233-243.
- Rice, J. A. and Wu, C. (2001) Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253-259.
- Riesz, F. and Nagy, B. S. (1990) *Functional Analysis*. New York: Dover Publications.
- Staniswalis, J. G. and Lee, J. J. (1998) Nonparametric regression analysis of longitudinal data. *J. Am. Statist. Ass.*, **93**, 1403-1418.

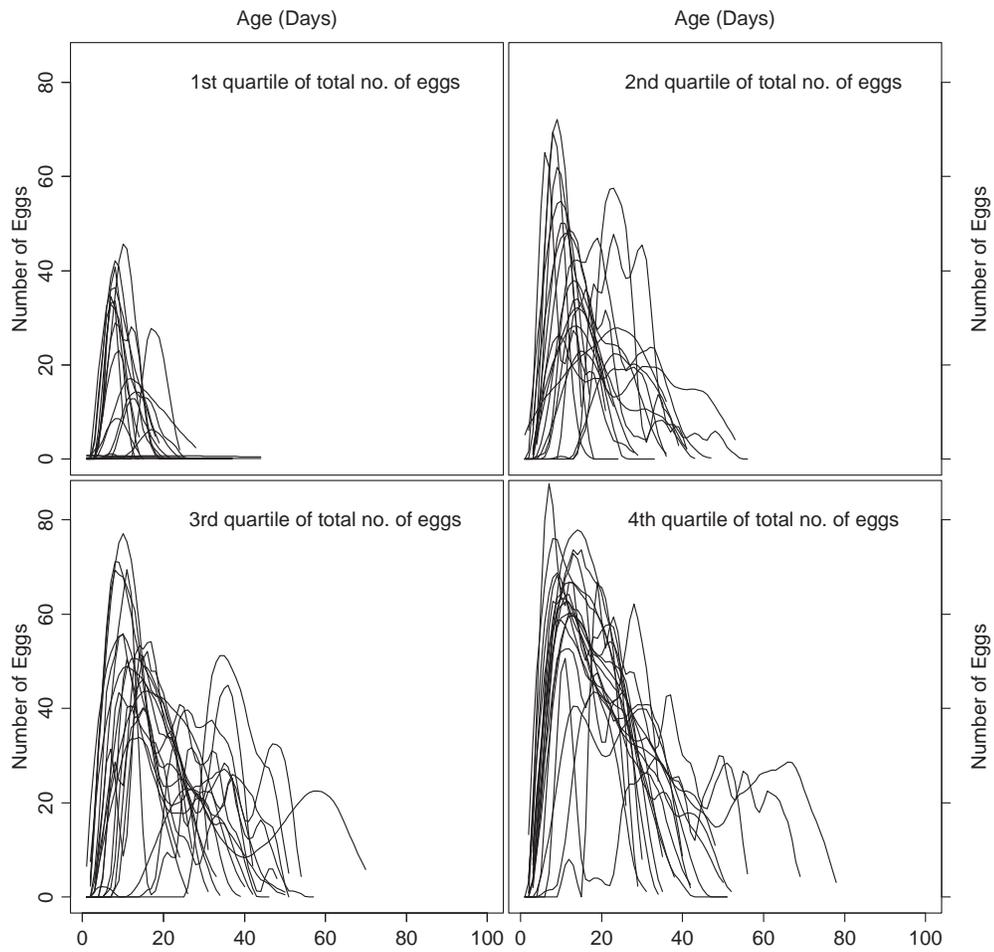


Figure 1: Twenty randomly selected smoothed egg-laying curves for each of the four quartiles of the total number of eggs.

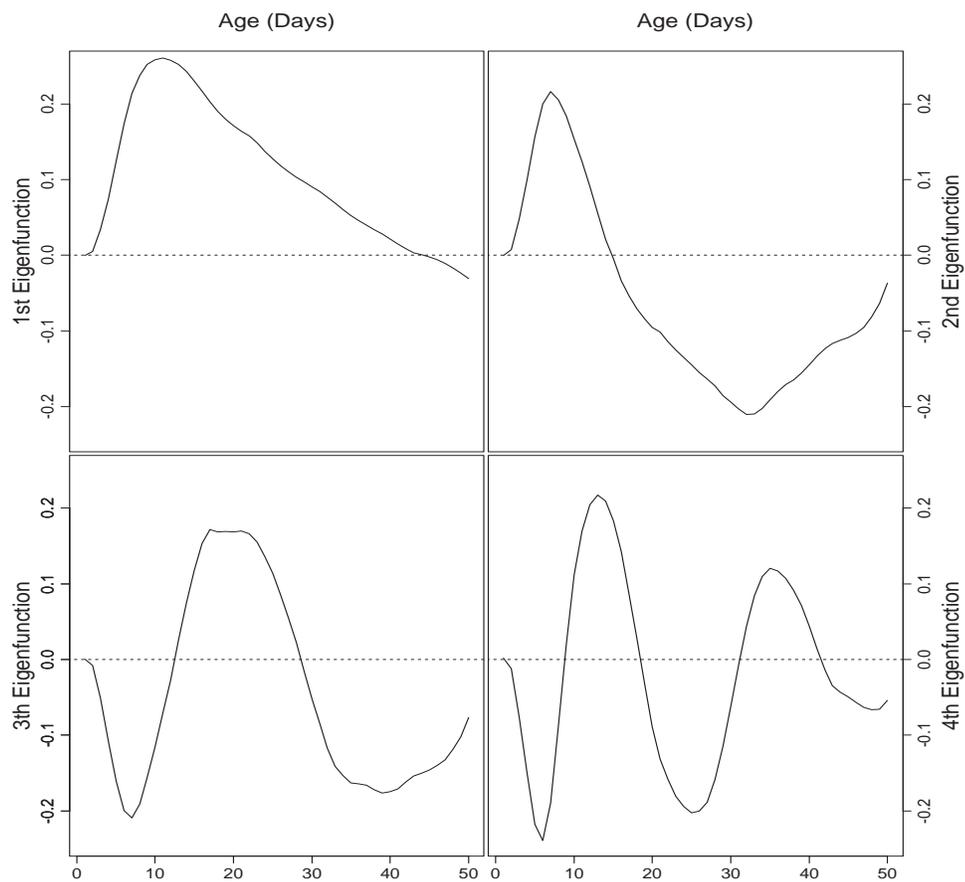


Figure 2: The first four eigenfunctions $\{\hat{\rho}_k\}_{k=1,\dots,4}$ (5) of the egg-laying data. The first eigenfunction explains 34.38%, the second additional 16.41%, the third additional 8.82%, and the fourth additional 4.93% of the total variance of the data. Bandwidths are selected by cross-validation individually by curve.

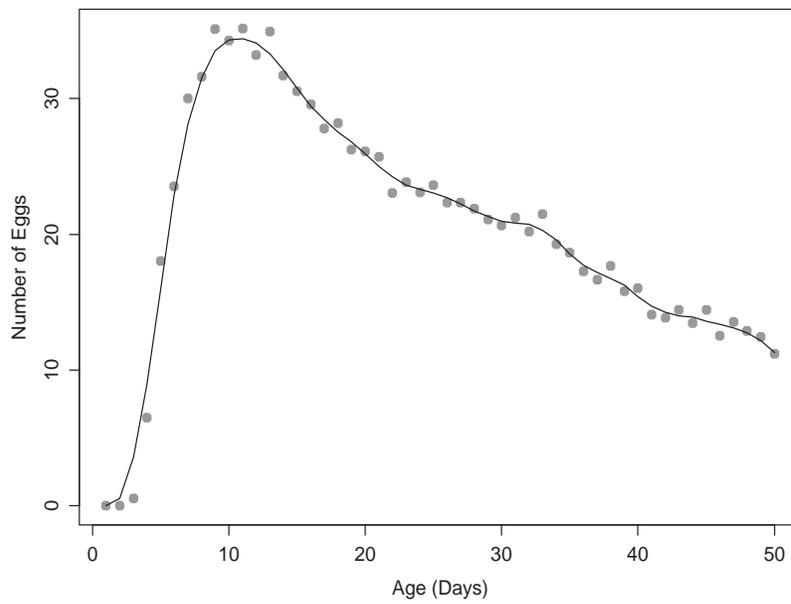


Figure 3: Estimated overall mean egg-laying function of time $\hat{\mu}(t)$ (3), with the leave-one-curve-out cross-validated bandwidth = 2.5 days.

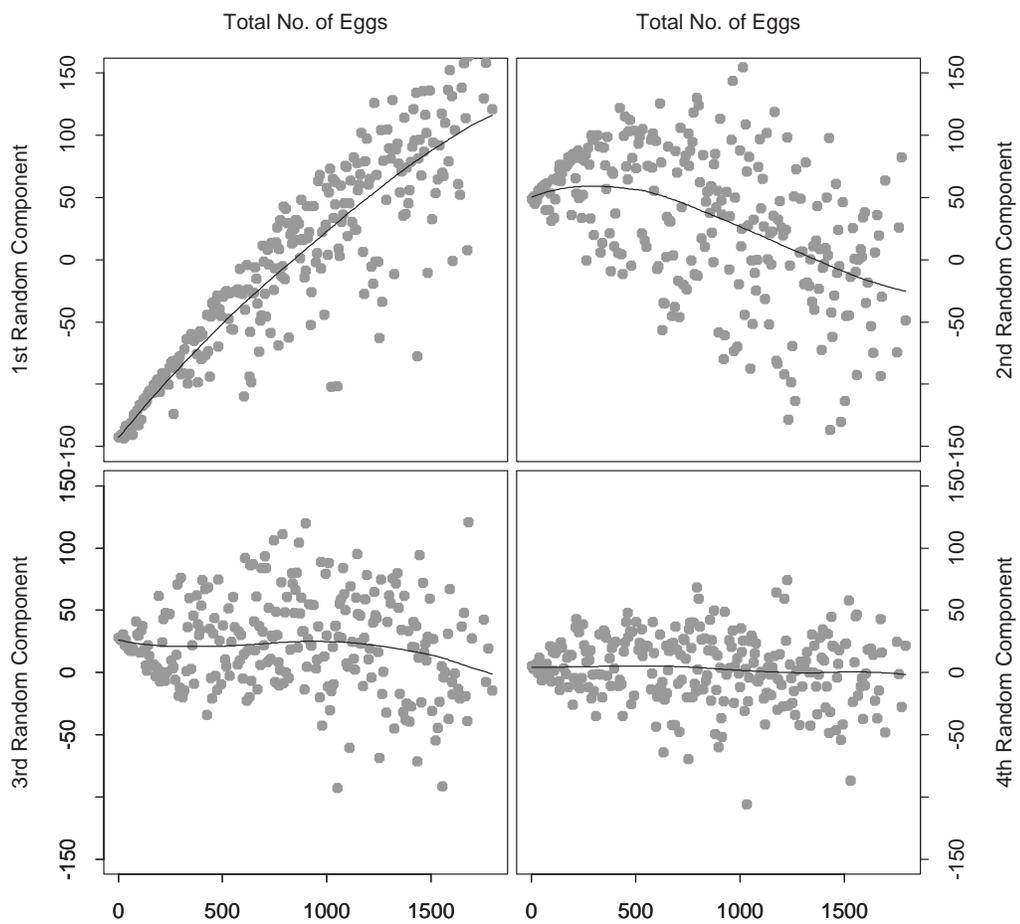


Figure 4: Smoothed random effects link functions $\{\hat{\alpha}_k(z)\}_{k=1,\dots,4}$ (8), in dependency on total number of eggs, based on a sample of 936 medflies. Bandwidths selected by cross-validation are 500 for all four curves.

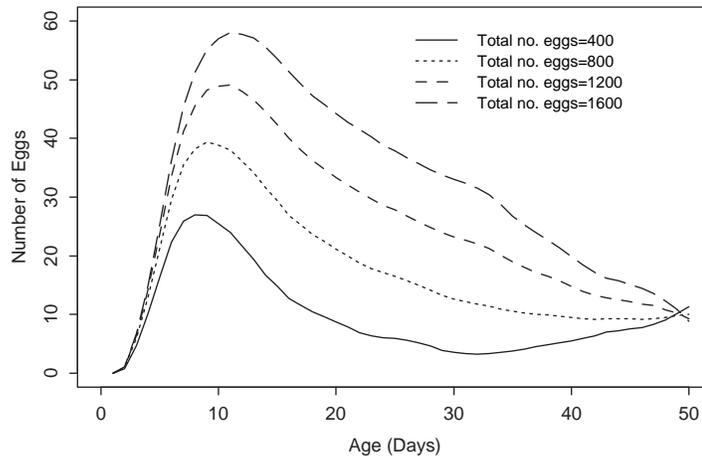
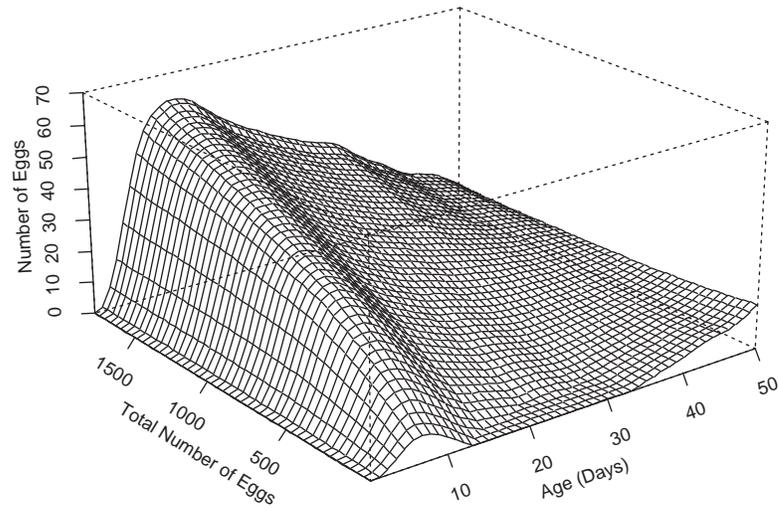


Figure 5: Upper panel: Fitted surface $\hat{\mu}_Z(t)$ (15) with total number of eggs as covariate. Lower panel: Cross-sections through the fitted surface $\hat{\mu}_Z(t)$, for total number of eggs fixed at 400, 800, 1200 and 1600.

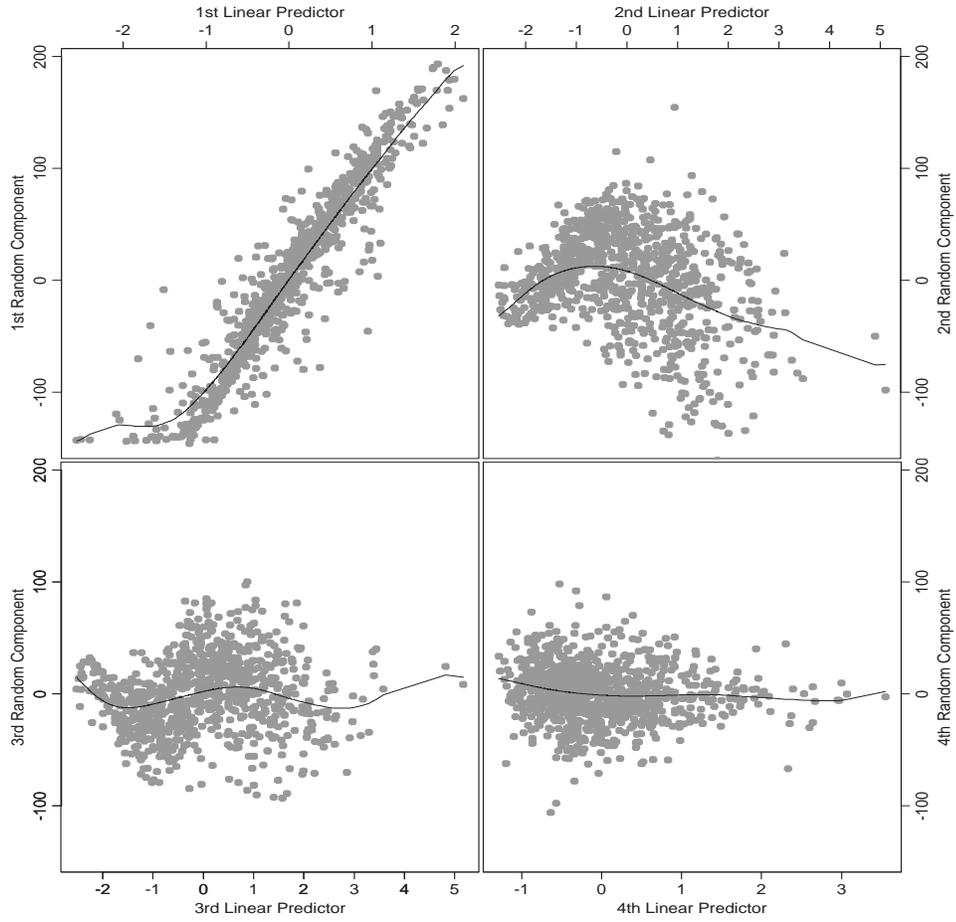


Figure 6: Estimated link functions for the random effects $\{\hat{\alpha}_k(\hat{\beta}'_k \mathbf{Z})\}_{k=1,\dots,4}$ (7), in dependency on the linear predictors, obtained the QLUE implementation of semi-parametric quasi-likelihood regression (SPQR).