

Time-Varying Functional Regression for Predicting Remaining Lifetime Distributions from Longitudinal Trajectories

Hans-Georg Müller* and Ying Zhang

Department of Statistics, University of California, One Shields Avenue,
Davis, California 95616, U.S.A.

*email: mueller@wald.ucdavis.edu

SUMMARY. A recurring objective in longitudinal studies on aging and longevity has been the investigation of the relationship between age-at-death and current values of a longitudinal covariate trajectory that quantifies reproductive or other behavioral activity. We propose a novel technique for predicting age-at-death distributions for situations where an entire covariate history is included in the predictor. The predictor trajectories up to current time are represented by time-varying functional principal component scores, which are continuously updated as time progresses and are considered to be time-varying predictor variables that are entered into a class of time-varying functional regression models that we propose. We demonstrate for biodemographic data how these methods can be applied to obtain predictions for age-at-death and estimates of remaining lifetime distributions, including estimates of quantiles and of prediction intervals for remaining lifetime. Estimates and predictions are obtained for individual subjects, based on their observed behavioral trajectories, and include a dimension-reduction step that is implemented by projecting on a single index. The proposed techniques are illustrated with data on longitudinal daily egg-laying for female medflies, predicting remaining lifetime and age-at-death distributions from individual event histories observed up to current time.

KEY WORDS: Aging; Biodemography; Conditional distribution; Dimension reduction; Event history analysis; Functional data analysis; Functional principal component; Longitudinal data; Prediction interval; Quantile estimation; Varying coefficient model.

1. Introduction

In longitudinal studies, frequently one obtains measurements on time-dependent covariates and a time-to-event for each individual. The association between these covariates and remaining survival is of interest in biodemography, where relationships between longevity and behavioral or reproductive longitudinal trajectories are studied. Typical biodemographic data obtained in studies on aging are uncensored and are obtained under controlled conditions with longitudinal covariates recorded on densely spaced regular time grids. Our methods are specifically addressing a need in biology and biodemography to analyze this type of data. A classical framework to characterize the relationship between longitudinal covariates and time-to-event would be to apply the proportional hazards regression model (Cox, 1972) with time-varying covariates. This model uses covariate information only at event times and determines the hazard rate at a given time by current covariate levels.

In contrast to the Cox model and similar approaches, we explore here models where remaining lifetime may depend on the entire event history as captured by the covariate trajectory, and not just on current covariate levels. That such an assumption is biologically reasonable was demonstrated, for example, in Müller et al. (2001), where a parametric model summarizing egg-laying trajectories of female medflies

(Mediterranean fruit fly, *Ceratitis capitata*) was shown to define remaining egg-laying potential. Thus a connection between the entire egg-laying trajectory up to current time and remaining lifetime was established. This model illustrated that age-at-death is related to early reproductive activity and shed new light on the cost of reproduction hypothesis (Partridge and Harvey, 1985). Issues in biodemographic hazard rate analysis and their implications were discussed in Vaupel et al. (1998), and are of ongoing interest in biodemography and the evolution of longevity (Oeppen and Vaupel, 2002).

In this article, we develop flexible functional models by using methods and tools from functional data analysis. We aim at nonparametric predictors of current remaining lifetime distributions, and especially mean remaining lifetime, by extracting information from the entire available covariate trajectory. For a situation where a sample of random trajectories is given with the same domain for each trajectory, commonly used functional data analysis techniques can be applied such as functional principal component analysis (Rice and Silverman, 1991) or functional regression (Cardot et al., 2003). Functional principal component analysis (FPCA) is an extension of multivariate PCA where one describes a sample of random trajectories by a mean function and the eigenfunctions of the covariance operator, also referred to as principal component

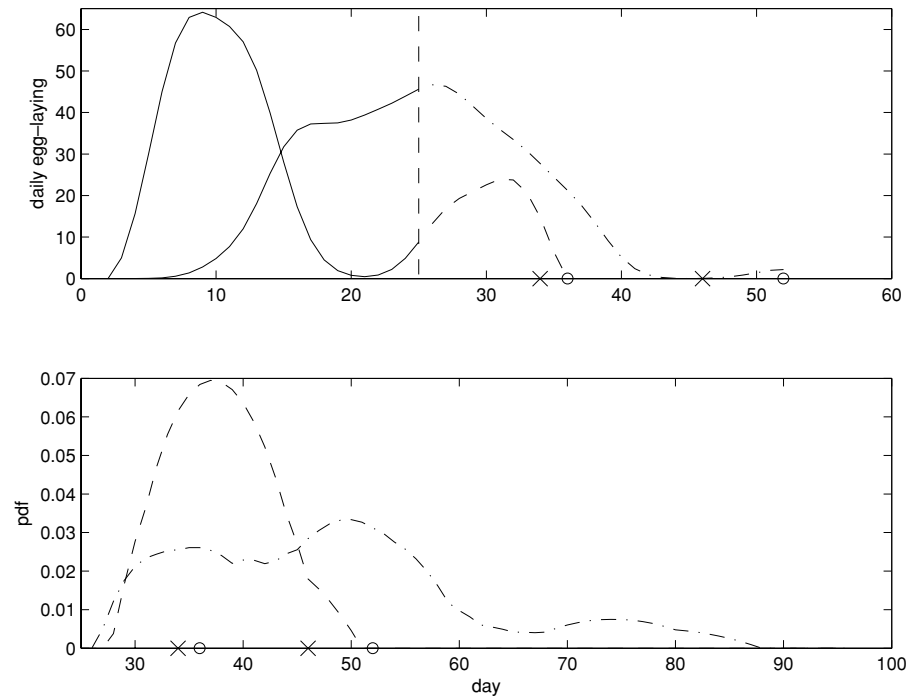


Figure 1. Smoothed daily egg-laying curves for two randomly selected medflies, from birth to death (upper panel). Data from birth to day 25 (solid curves in upper panel) are used for predicting density functions of remaining lifetime. Actually observed age-at-death is marked as a circle, predicted age-at-death as a cross for each fly. Predicted densities of remaining lifetime for each of the two flies are shown in the lower panel, with line patterns matched to the egg-laying curves in the upper panel.

functions. These provide a parsimonious orthonormal basis in which to represent the observed trajectories and correspond to the “modes of variation” of random trajectories (Castro, Lawton, and Sylvestre, 1987). The coefficients of the eigenfunctions in an eigenbasis representation of a given random trajectory are random variables, and a finite number of these are commonly used as predictors for functional regression models.

We propose two main extensions of this methodology: first, to the case of a response that is a distribution function rather than a number. This will allow the estimation of quantiles and prediction intervals for remaining lifetimes, which are highly desirable for survival outcomes. We will introduce appropriate assumptions in Section 2.3 and discuss estimation of remaining lifetime distributions in Section 3.2. Second, in the situation we consider, the predictor trajectories for each individual are only observed until age-at-death, and therefore, do not share a common domain. At each given time only a random number of subjects is still alive for whom the covariate trajectories can be observed. Accordingly, any postulated functional regression model must evolve over time. This motivates the time-varying functional regression models that we introduce in Section 2.2, with appropriate estimates being discussed in Section 3.1.

Besides the practical purpose of prediction, it is usually of scientific interest to determine which features of a covariate history are related to survival and longevity. A related area is joint modeling (Self and Pawitan, 1992; Tsiatis, DeGruttola, and Wulfsohn, 1995; Bycott and Taylor, 1998), where statis-

tical models reflect the impact of current covariate values on hazard ratio and current mortality, while our goal is to predict the distribution of remaining lifetimes for each subject still alive at age t , and also to address the simpler problem of predicting mean remaining lifetime, given the entire event history information for a subject up to current time t . As an illustration, the upper panel of Figure 1 provides the egg-laying curves for two randomly selected medflies, with respective age-at-death of 36 and 52 days. Assuming that only the solid segments of the covariate processes are observed from birth until age $t = 25$ days, these observed trajectories then form the basis for these predictions.

The article is organized as follows: In Section 2, we introduce assumptions and the time-varying functional regression models that result from coupling time-varying functional principal components with ideas from functional regression. In Section 3, we describe estimation of the components of the model. Application of the proposed model to a biodemographic study, using egg-laying trajectories of female medflies as predictors, and comparisons of the performance with other possible estimators are the theme of Section 4, followed by discussion and concluding remarks in Section 5.

2. Proposed Model

2.1 Modeling Predictor Trajectories

We denote lifetime or age-at-death of a subject by T , and assume that for each subject a covariate X is recorded continuously or on a dense regular grid until death. Then $\{X(t), T\}$, $t \in [0, T]$ are the data available for each subject, where $X(\cdot)$

is the covariate trajectory, with support $[0, T]$ determined by the lifetime T . We denote the observed covariate trajectory for an individual that is still alive at time t by $\tilde{X}(s, t)$, $0 \leq s \leq t$, that is,

$$\tilde{X}(s, t) = \begin{cases} X(s) & \text{if } T \geq t, 0 \leq s \leq t, \\ \text{unobserved} & \text{if } T < t. \end{cases}$$

Our aim is to predict the distribution of remaining lifetime $T - t$, given $\tilde{X}(s, t)$, $0 \leq s \leq t$, and in particular mean remaining lifetime.

For $t \in [0, \tau]$, with some large τ , define

$$\mu(s, t) = E\{X(s) | T > t\}, \quad (1)$$

and the eigenfunctions or principal component functions of the conditional covariances as solutions of the eigenequations

$$\int_0^t \text{cov}\{X(s), X(u) | T > t\} \rho_{jt}(u) du = \lambda_{jt} \rho_{jt}(s), \quad (2)$$

where $\lambda_{1t} \geq \lambda_{2t} \geq \dots > 0$ are eigenvalues and $\rho_{1t}(\cdot), \rho_{2t}(\cdot), \dots$, are orthonormal eigenfunctions associated with these eigenvalues. Then one has the representation, for $0 \leq s_1, s_2 \leq t$,

$$\begin{aligned} \text{cov}\{X(s_1), X(s_2) | T > t\} \\ = \sum_{j=1}^{\infty} \lambda_{jt} \rho_{jt}(s_1) \rho_{jt}(s_2), \quad 0 \leq s_1, s_2 \leq t. \end{aligned} \quad (3)$$

We find that

- (i) $E\{\tilde{X}(s, t)\} = \mu(s, t) = E\{X(s) | T > t\}$, and
- (ii) $\text{cov}\{\tilde{X}(s_1, t), \tilde{X}(s_2, t)\} = \text{cov}\{X(s_1), X(s_2) | T > t\}$.

As one can easily show, the trajectories observed conditional on survival beyond t are independent, and therefore, the observed trajectories $\tilde{X}(s, t)$ for subjects with $T > t$ can be represented by the Karhunen–Loève expansion (Ash and Gardner, 1975; Rice and Silverman, 1991),

$$\tilde{X}(s, t) = \mu(s, t) + \sum_{j=1}^{\infty} \varepsilon_{jt} \rho_{jt}(s), \quad 0 \leq s \leq t, \quad t \geq 0, \quad (4)$$

where the random variables ε_{jt} are conditional functional principal component scores with $E(\varepsilon_{jt}) = 0$, $\text{var}(\varepsilon_{jt}) = \lambda_{jt}$, and $\text{cov}(\varepsilon_{jt}, \varepsilon_{kt}) = 0$ (for $j \neq k$). Here $\tilde{X}(s, t)$ are trajectories that are observed on the entire interval $[0, t]$, and the functional principal component scores can be represented as

$$\varepsilon_{jt} = \int_0^t \{\tilde{X}(s, t) - \mu(s, t)\} \rho_{jt}(s) ds. \quad (5)$$

The number N_t of trajectories $\tilde{X}(s, t)$ observable up to time t is random. Assuming $\bar{F}(t) = P(T > t)$, we have $N_t \sim \text{Binomial}(n, \bar{F}(t))$, where n is the total number of subjects. Denoting by $R(t)$ the risk set at time t , $R(t) = \{i : T_i > t\}$, then $\tilde{X}_i(s, t) \stackrel{\text{i.i.d.}}{\sim} \tilde{X}(s, t)$ for all $i \in R(t)$.

The remaining lifetime function at t is

$$r(t) = E(T - t | 0 \leq t \leq T), \quad (6)$$

and the distribution function of remaining lifetime at y , where $y \geq 0$, is

$$F_t(y) = P(T - t \leq y | 0 \leq t \leq T), \quad (7)$$

so that $r(t) = \int_t^{\infty} y dF_t(y)$. It is well known (Cox, 1972) that the corresponding survival $\bar{F}(\cdot)$ and hazard $\lambda(\cdot)$ functions are $\bar{F}(t) = \frac{r(0)}{r(t)} \exp\{-\int_0^t \frac{1}{r(u)} du\}$ and $\lambda(t) = \{\frac{d}{dt} r(t) + 1\}/r(t)$, so that the function $r(\cdot)$ defines the survival schedule.

2.2 Modeling Mean Remaining Lifetime

Our aim is to relate the remaining lifetime $T - t$ for a given arbitrary $t \in [0, \tau]$ to the observed trajectory \tilde{X} on $[0, t]$, that is, to estimate

$$r_{\tilde{X}}(t) = E\{T - t | \tilde{X}(s, t), \quad 0 \leq s \leq t, \quad 0 \leq t \leq T\}. \quad (8)$$

We assume that there exists a family of smooth link functions h_t indexed by t , with $h_t(s) = H(t, s) : [0, \tau] \times \mathbf{R} \rightarrow \mathbf{R}$, for a function H that is continuous in s and t , and an associated evaluation function $\beta(s, t)$, satisfying $\beta \in L^2(C_t)$, $C_t = \{(s, t), 0 \leq s \leq t, 0 \leq t \leq \tau\}$, such that

$$r_{\tilde{X}}(t) = h_t \left\{ \int_0^t \tilde{X}(s, t) \beta(s, t) ds \right\}. \quad (9)$$

This assumption puts our model into the framework of an extension of functional regression (Ramsay and Silverman, 1997; Cardot et al., 2003; Müller, 2005), where the extension is the inclusion of a time-varying feature, while the classical linear functional regression model would be $E(y | X) = \int_0^S X(s) \beta(s) ds$ with a fixed $S > 0$, with generalized version $E(y | X) = g\{\int_0^S X(s) \beta(s) ds\}$, where g is a link function (James, 2002; Müller and Stadtmüller, 2005). Another type of extension of functional regression where current observed values are to be predicted from past observations was proposed in Malfait and Ramsay (2003).

For given t and any orthonormal basis, $\psi_{jt}(\cdot)$, $j = 1, 2, \dots$, on $L^2([0, t])$, the evaluation function $\beta(\cdot, t)$ can be represented by $\beta(s, t) = \sum_{j=1}^{\infty} \beta_{jt} \psi_{jt}(s)$, $0 \leq s \leq t, 0 \leq t \leq \tau$, with varying coefficients β_{jt} . A special choice for the basis are the eigenfunctions $\rho_{jt}(\cdot)$ of $\text{cov}(X(u), X(v))$. Then the model for mean remaining lifetime becomes

$$r_{\tilde{X}}(t) = h_t \left[r_0(t) + \int_0^t \{\tilde{X}(s, t) - \mu(s, t)\} \beta(s, t) ds \right], \quad (10)$$

and because $r_0(t)$ is a nonrandom function, we may introduce another (nonrandom) smooth link function $g_t(\cdot)$ such that $g_t\{z(t)\} = h_t\{r_0(t) + z(t)\}$ for an arbitrary function $z(t)$. This leads to

$$r_{\tilde{X}}(t) = g_t \left[\int_0^t \{\tilde{X}(s, t) - \mu(s, t)\} \beta(s, t) ds \right], \quad (11)$$

where $\int_0^t \{\tilde{X}(s, t) - \mu(s, t)\} \beta(s, t) ds$ plays the same role as a linear predictor in a generalized linear model, with the additional feature that it is a function of t , and therefore, is referred to as the linear predictor function. If $\tilde{X}(\cdot, t)$ and $\beta(\cdot, t)$ are expressed in terms of the orthonormal basis $\rho_{1t}(\cdot), \rho_{2t}(\cdot), \dots$, then by the orthogonality of the basis functions, the linear predictor function becomes $\int_0^t \{\tilde{X}(s, t) - \mu(s, t)\} \beta(s, t) ds = \sum_{j=1}^{\infty} \varepsilon_{jt} \beta_{jt}$, where ε_{jt} are as in (5) and

$$\beta_{jt} = \int_0^t \beta(s, t) \rho_j(s) ds. \quad (12)$$

We assume there exists a finite number M of components (where $M = M(n)$ may increase with sample size) such that processes $\tilde{X}(\cdot, \cdot)$ are sufficiently well approximated by projecting on the function space spanned by the first M eigenfunctions. It then suffices to consider truncated linear predictor functions given by

$$\eta(t) = \sum_{j=1}^M \varepsilon_{jt} \beta_{jt}. \quad (13)$$

These developments suggest a varying coefficient generalized linear model for remaining lifetime, with the time-varying principal component scores ε_{jt} as predictors. Once a number M of component scores to be included in the model as predictors has been chosen, the trajectory $\tilde{X}(\cdot, t)$ is summarized by the random M -vector $(\varepsilon_{1t}, \dots, \varepsilon_{Mt})$. We then fit the varying coefficient generalized linear model (Hastie and Tibshirani, 1993)

$$r_{\tilde{X}}(t) = g_t \left(\sum_{j=1}^M \varepsilon_{jt} \beta_{jt} \right), \quad (14)$$

where $g_t(\cdot)$ is the link function at time t which may be assumed to be known or unknown but smooth. For quasi-likelihood-based estimating equations, we also need to specify a variance function, $V(\cdot, \cdot)$,

$$\begin{aligned} V(t, \tilde{X}) &= \text{var}\{T - t \mid \tilde{X}(s, t), \quad 0 \leq s \leq t, \quad 0 \leq t \leq T\} \\ &= \text{var}(T - t \mid \varepsilon_{1t}, \dots, \varepsilon_{Mt}) = \sigma^2\{r_{\tilde{X}}(t)\}, \end{aligned} \quad (15)$$

which depends on the conditional mean $r_{\tilde{X}}(t)$ only via the smooth function $\sigma^2(\cdot)$. This function also can be assumed to be known (such as in quasi-Poisson or quasi-Gamma regression) or unknown but smooth.

When the link function $g_t(\cdot)$ is chosen as identity function for all t , model (14) is replaced by a varying coefficient linear regression model,

$$r_{\tilde{X}}(t) = r_0(t) + \sum_{j=1}^M \varepsilon_{jt} \beta_{jt}, \quad (16)$$

where $r_0(t)$ is the mean remaining lifetime function at t , corresponding to a varying intercept function, ε_{jt} are the individual random components (uncorrelated and with zero mean), and $\beta_{jt}, j = 1, 2, \dots, M$ are the varying coefficients.

2.3 Modeling Conditional Distributions of Remaining Lifetime

The conditional distribution of remaining lifetime $T - t$, at a current time (age) t of a subject, and given an observed trajectory \tilde{X} up to time t , is defined as $F_{\tilde{X},t}(y) = P\{T - t \leq y \mid \tilde{X}(s, t), 0 \leq s \leq t, t \leq T\}$. Once these conditional distributions have been determined, other quantities that are functionals of conditional distribution functions such as conditional quantiles and conditional densities can be constructed as well.

Having summarized the covariate trajectory $\tilde{X}(\cdot, t)$ by the linear predictor function $\eta(\cdot)$ (13), we add the assumption that the linear predictor function determines the conditional distribution,

$$\begin{aligned} F_{\tilde{X},t}(y) &= P\{T - t \leq y \mid \tilde{X}(s, t), 0 \leq s \leq t, t \leq T\} \\ &= P\{T - t \leq y \mid \eta(t)\} = \phi_{t,y}\{\eta(t)\}, \end{aligned} \quad (17)$$

where the unknown functions $\phi_{t,y}(\cdot)$ are assumed to be smooth in y and t and in their argument. Estimating conditional remaining lifetime distributions then is equivalent to estimating functions $\phi_{t,y}(\cdot)$ and $\eta(\cdot)$. By substituting the linear predictor function $\eta(t)$ in lieu of the observed event history process to time t , $\tilde{X}(s, t)$, $0 \leq s \leq t$, we have reduced the initially infinite-dimensional predictor to dimension one. This corresponds to a major dimension reduction step. We assume that $\phi_{t,y}(\cdot)$ is smooth or may alternatively choose a parametric model for $\phi_{t,y}(\cdot)$. We obtain $P(T - t \leq y \mid \eta)$ by estimating the functions $\phi_{t,y}(\cdot)$ nonparametrically, as detailed in Section 3.3.

3. Estimating the Model Components

3.1 Preliminaries

For smoothing purposes, we find it convenient to use local linear scatterplot smoothers for the nonparametric estimation of a regression function $E(y \mid X = x)$. We note that many other smoothing methods could be used, for example, various spline smoothers. Given data $\{(x_i, y_i) \in \mathbf{R}^2, i = 1, \dots, n\}$, these are implemented by weighted local least squares, where

$$\sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \{y_i - b_0 - b_1(x - x_i)\}^2 \quad (18)$$

is minimized with respect to (b_0, b_1) . Here, $K \geq 0$ is a non-negative kernel function, chosen as a probability density, and h a suitably chosen bandwidth (smoothing parameter). The resulting nonparametric regression estimate is

$$\begin{aligned} S\{x; (x_i, y_i)_{i=1, \dots, n}; h\} &= \hat{E}\{y \mid X = x\} \\ &= \hat{b}_0 = \frac{\sum_{i=1}^n w_i(x, h) y_i}{\sum_{i=1}^n w_i(x, h)}, \end{aligned} \quad (19)$$

which is linear in the data y_i , with weight functions $w_i(x, h)$ given by

$$\begin{aligned} w_i(\mathbf{x}, h) &= \frac{1}{h} K\left(\frac{x - x_i}{h}\right) \\ &\times \left\{ \frac{1}{h} \sum_j K\left(\frac{x - x_j}{h}\right) (x - x_j)^2 \right. \\ &\quad \left. - (x - x_i) \frac{1}{h} \sum_j K\left(\frac{x - x_j}{h}\right) (x - x_j) \right\} \end{aligned} \quad (20)$$

(see, e.g., Fan and Gijbels, 1996).

A first step is to estimate the mean function $\mu(\cdot, \cdot)$ (1) at current time t for all subjects who are at risk at t , i.e., for which $T > t$, where T is age-at-death. The estimate of $\mu(s, t)$ is

$$\hat{\mu}(s, t) = S[s; \{t_{ij}, \tilde{X}_i(t_{ij}, t)\}_{i \in R(t), j \in \{j: 0 \leq t_{ij} \leq t\}}; h], \quad (21)$$

where $R(t)$ is the risk set at time t , the t_{ij} 's are the pooled time points of all observations at which the i th subject was

observed prior to time t , provided that $T_i > t$, and h is the smoothing parameter of the scatterplot smoother S , which is evaluated at the argument s ($0 \leq s \leq t$). The number M of components that are included in model (14) or (16) and the necessary bandwidths can be determined by one-curve-leave-out cross-validation (Rice and Silverman, 1991) or according to the amount of variation explained by the first M components.

We use two-dimensional local linear smoothers to obtain the covariance surface of observed processes up to time t and then estimate the eigenfunctions and eigenvalues. Additional measurement errors may contaminate the observed covariates, i.e., the observed data are $(\tilde{X}_i, T_i), i = 1, \dots, n$, where T_i is observed age-at-death for the i th subject and observed covariates are

$$\tilde{X}_i(t_{ij}, t) = \tilde{X}(t_{ij}, t) + \epsilon_{ij}, \quad i \in R(t), 0 \leq t_{ij} \leq t, \quad (22)$$

with $E(\epsilon_{ij}) = 0$, $\text{var}(\epsilon_{ij}) = \sigma^2$, and $\text{cov}(\epsilon_{ij}, \epsilon_{il}) = 0$ for any $j \neq l$. In this case,

$$\begin{aligned} \text{cov}\{\tilde{X}_i(t_{ij}, t), \tilde{X}_i(t_{il}, t)\} \\ = \text{cov}\{\tilde{X}(t_{ij}, t), \tilde{X}(t_{il}, t)\} + \sigma^2 \delta_{jl}, \end{aligned} \quad (23)$$

where δ_{jl} is 1 if $j = l$ and 0 otherwise. The raw covariances $G_i^t(t_{ij}, t_{il}) = \{\tilde{X}_i(t_{ij}, t) - \hat{\mu}(t_{ij}, t)\}\{\tilde{X}_i(t_{il}, t) - \hat{\mu}(t_{il}, t)\}, i \in R(t), 0 \leq t_{ij}, t_{il} \leq t$, with $\hat{\mu}$ from (21), contain the measurement errors in the diagonal, therefore unrestricted 2-dimensional smoothing is not feasible, as it would produce biased covariance estimates in the neighborhood of the diagonal. We, therefore, apply a 2-dimensional scatterplot smoother to off-diagonal elements only, to obtain the smoothed covariance surface denoted as $\hat{G}^t(r, s)$. Details can be found in Yao et al. (2003); compare also Staniswalis and Lee (1998) for similar considerations.

Estimated eigenfunctions and eigenvalues are solutions of estimated eigenequations,

$$\int_0^t \hat{G}^t(r, s) \rho_{jt}(r) dr = \lambda_{jt} \rho_{jt}(s), \quad (24)$$

where eigenfunctions are subject to the constraints $\int_0^t \rho_{jt}(s)^2 ds = 1$ and $\int_0^t \rho_{jt}(s) \rho_{lt}(s) ds = 0$ for $j < l$. These solutions are found by discretizing (24), obtaining corresponding discrete solutions, and then smoothing them to obtain eigenfunction estimates; appropriate modifications as described in Yao et al. (2003) are made in case \hat{G}^t is not positive definite. The functional principal component scores are then determined by

$$\hat{\epsilon}_{ijt} = \int_0^t \{\tilde{X}_i(s, t) - \hat{\mu}(s, t)\} \hat{\rho}_{jt}(s) ds, \quad (25)$$

which can be obtained by numerical integration or, in the Gaussian case, by a conditioning argument. Consistency results for $\hat{\rho}_{jt}, \hat{\mu}$, and $\hat{\epsilon}_{ijt}$ can be found in an unpublished manuscript by P. Hall and M. Hosseini-Nasab (2003).

3.2 Estimating Mean Remaining Lifetime

Denoting the least squares estimates of $\beta_t = (\beta_{0t}, \dots, \beta_{Mt})$ in model (16) by $\hat{\beta}_t$,

$$\hat{\beta}_t = \underset{\beta_t}{\text{argmin}} \sum_{i \in R(t)} \left\{ T_i - t - \left(\beta_{0t} + \sum_{j=1}^M \hat{\epsilon}_{ijt} \beta_{jt} \right) \right\}^2, \quad (26)$$

the fitted model for remaining lifetime is

$$\hat{r}_{\tilde{X}_i}(t) = \hat{\beta}_{0t} + \sum_{j=1}^M \hat{\epsilon}_{ijt} \hat{\beta}_{jt}, \quad (27)$$

where $\hat{\beta}_{0t}$ is the mean remaining lifetime function. In some cases (as it turns out, not in our application example), smoothness in t for varying coefficients β_{jt} may be exploited to improve the varying coefficient estimates by a smoothing step (compare, e.g., Cai, Fan, and Li [2000] or Wu and Yu [2002] for smoothing varying coefficients). We require that the eigenfunctions $\rho_{jt}(s)$ are continuous in a suitable function norm as t varies. The sign of the eigenfunctions is generally undetermined. If t is increased by a small $\delta > 0$, we choose the sign of the eigenfunction at $t + \delta$ by determining

$$\hat{\zeta} = \underset{\zeta \in \{-1, 1\}}{\text{argmin}} \int_0^t \{\zeta \hat{\rho}_{j(t+\delta)}(s) - \hat{\rho}_{jt}(s)\}^2 ds, \quad (28)$$

and declaring $\hat{\zeta} \hat{\rho}_{j(t+\delta)}(s)$ to be the j th eigenfunction at time $t + \delta$.

For the fitted model (27), the one-leave-out prediction for the i th subject is

$$\hat{r}_{\tilde{X}_i}^{(-i)}(t) = \hat{r}_{0t}^{(-i)} + \sum_{j=1}^M \tilde{\epsilon}_{ijt} \hat{\beta}_{jt}^{(-i)}, \quad (29)$$

where $\tilde{\epsilon}_{ijt}$ are the coefficients for the i th subject obtained for eigenfunctions $\hat{\rho}_j^{(-i)}(t), j = 1, \dots, M$, which are the estimated eigenfunctions after removing the i th subject's trajectory. The one-leave-out predictions lead to the root squared prediction errors at t ,

$$\text{RSPE}(t) = \left[\frac{1}{N_t} \sum_{i \in R(t)} \left\{ \hat{r}_{\tilde{X}_i}^{(-i)}(t) - (T_i - t) \right\}^2 \right]^{\frac{1}{2}}, \quad (30)$$

where N_t is the number of subjects in the risk set $R(t)$. Root squared prediction error functions for various predictors that are described in Section 4 are displayed below in Figure 5.

For generalized linear regression models, Chiou and Müller (1998, 1999) developed a nonparametric quasi-likelihood with unknown link and variance function estimation (QLUE) algorithm by substituting nonparametric estimates for link and variance functions in lieu of the true link and variance functions in the usual definition of quasi-likelihood. The estimating procedure is based on a three-stage iteration, which consists of three updating steps, nonparametric updating of the link function, nonparametric updating of the variance function, and updating of the regression parameters. Adapting to the current situation, and setting $\hat{\epsilon}_{it} = (\hat{\epsilon}_{i1t}, \dots, \hat{\epsilon}_{ipt})^T$, one aims at solving the semiparametric estimating equation

$$U^*(\beta_t) = \sum_{i=1}^{N_t} \frac{T_i - \hat{g}\{\tilde{\eta}_i(t); \beta_t\}}{\hat{\sigma}^2[\hat{g}\{\tilde{\eta}_i(t); \beta_t\}]} \cdot \hat{g}_1\{\tilde{\eta}_i(t); \beta_t\} \hat{\epsilon}_{it} = 0 \quad (31)$$

with respect to β_t . Here $\tilde{\eta}_i(t) = \sum_{j=1}^M \tilde{\epsilon}_{ijt} \beta_{jt}$ is the linear predictor function with estimated predictors $\tilde{\epsilon}_{ijt}$ (25), $\hat{g}(\cdot; \hat{\beta}_t)$

and $\hat{g}_1(\cdot; \hat{\beta}_t)$ are nonparametric smooth estimates of the link function, respectively, its first derivative, and $\hat{\sigma}^2(\cdot)$ is a smooth estimate of the variance function.

When applying QLUE, we use data-based bandwidth selection for smoothing link and variance functions based on Pearson's chi-square by obtaining

$$\operatorname{argmin}_{b_0, b} \left| \sum_{i=1}^{N_t} \frac{(T_i - \hat{\mu}_{b_0 i})^2}{\hat{\sigma}_b^2(\hat{\mu}_{b_0 i})} - (N_t - p) \right|, \quad (32)$$

where $\hat{\mu}_{b_0 i}$ denotes the estimated predictors based on the link function obtained with bandwidth b_0 and $\hat{\sigma}_b^2(\cdot)$ denotes the nonparametric variance function estimate obtained with the bandwidth b , where p is the number of predictors, $p = M$ in our case. For details see Chiou and Müller (1998). The QLUE algorithm can be shown to provide consistent estimates $\hat{\beta}_t, \hat{g}(\cdot)$, and $\hat{\sigma}^2(\cdot)$ for the components of models (14) and (15), for each fixed t , given predictors ε_{ijt} , and this result can be extended to the case where true predictors ε_{ijt} are replaced by consistent estimates $\hat{\varepsilon}_{ijt}$. The resulting estimate of the linear predictor function $\eta(t)$ (13) is

$$\hat{\eta}(t) = \sum_{j=1}^M \hat{\varepsilon}_{jt} \hat{\beta}_{jt}. \quad (33)$$

3.3 Estimating Conditional Distributions

Our approach to conditional distribution estimation is based on the single index assumption (17), which will provide the necessary dimension reduction, $P\{T - t \leq y | \tilde{X}(s, t), 0 \leq s \leq t, t \leq T\} = P\{T - t \leq y | \eta(t)\}$. Here $\eta(t) = \sum_{j=1}^M \varepsilon_{jt} \beta_{jt}$ is the single index or linear predictor function which can be estimated via (31).

We then obtain nonparametric smooth estimates of the conditional distribution with relative ease: Consider i.i.d.

where $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$ for a bandwidth h . Combining estimate (34) with (17),

$$\begin{aligned} \hat{F}\{y | \tilde{X}(s, t), 0 \leq s \leq t, t \leq T\} \\ &= \hat{P}\{T - t \leq y | \eta(t) = \eta\} \\ &= \frac{\sum_{i \in \{i: T_i - t \leq y\} \cap R(t)} K_h\{\eta - \hat{\eta}_i(t)\}}{\sum_{j \in R(t)} K_h\{\eta - \hat{\eta}_j(t)\}}, \end{aligned} \quad (35)$$

where $\hat{\eta}_i(t)$ (33) is the linear predictor for the i th individual and $R(t)$ the risk set at time t .

We can also apply a conditional density estimator (see, e.g., Yu and Jones, 1998; Hall and Müller, 2003), extending the notion of a conditional density to functional predictors. Assuming h_y is the bandwidth for the density estimation step, K_y is the corresponding kernel, and h and K are bandwidth and kernel for the predictor η , respectively, we may estimate the conditional density $f(y | \varepsilon_{jt}, j = 1, \dots, M) = f_t(y | \eta(t) = \eta)$ by

$$\begin{aligned} \hat{f}_t(y | \eta) &= \sum_{i \in \{j: T_j - t \leq y\} \cap R(t)} \frac{1}{h_y} K_y \left\{ \frac{y - (T_i - t)}{h_y} \right\} \\ &\times \frac{K_h\{\eta - \hat{\eta}_i(t)\}}{\sum_{j \in R(t)} K_h\{\eta - \hat{\eta}_j(t)\}}. \end{aligned} \quad (36)$$

Analogously, we obtain estimates for conditional quantile and quantile density functions.

In order to avoid boundary bias, we implement (36) with boundary kernels (Müller and Wang, 1994). For example, for arguments y in the left boundary region, $0 \leq y \leq h_y$, we define $q = \frac{y}{h_y}$, and when using Epanechnikov kernels for estimating the conditional density, the kernels in the interior and at the boundary may be chosen as

$$K_y(x) = \begin{cases} K_0 \equiv \frac{3}{4}(1 - x^2)_{[-1, 1]}, & \text{for } y \geq h_y \\ K_q \equiv 3 \left(\frac{1}{1 + q} \right)^3 (x + 1) \left[(3q - 1) + \frac{3(1 - q)^2}{1 + q} + x \left\{ \frac{6(1 - q)}{1 + q} - 2 \right\} \right]_{[-1, q]}, & \text{for } 0 \leq y < h_y. \end{cases}$$

pairs, $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$, where $(X, Y) \in \mathbf{R}^2$. Then to estimate the conditional distribution function $F(y | x) = P(Y \leq y | X = x)$ from this sample, note that $F(y | x) = E\{I(Y_i \leq y) | X = x\}$, where I is the indicator function. Estimation of the conditional distribution function can thus be framed as a regression problem. Applying nonparametric regression, for example, via Nadaraya-Watson kernel estimators,

$$\hat{F}(y | X = x) = \frac{\sum_{i: Y_i \leq y} K_h(x - X_i)}{\sum_{i=1}^n K_h(x - X_i)}, \quad (34)$$

If we use local polynomial smoothers, these will automatically include boundary adjustments through "equivalent" boundary kernels.

3.4 Inference via Bootstrap

We propose to use bootstrapping to obtain inference for coefficients β_{jt} (12), based on estimates derived for either the linear model (16) or the generalized linear model (14). To generate bootstrap samples, we sample n times with replacement from the set of all subjects (\mathbf{X}_i, T_i) , $i = 1, \dots, n$. For each sampled subject we include its observed covariate trajectory and the associated age-at-death.

For each bootstrap sample, we then estimate the coefficients β_{jt} at each time point t , using the method of choice such as (26) or (31). Obtaining the proposed estimates for

each of the bootstrap samples, we end up with B bootstrap estimates β_{jt}^* . Finally, from the resulting bootstrap distributions F_t^* we may construct pointwise confidence intervals for each t at level α , by finding the empirical percentiles at levels $\alpha/2$ and $(1 - \alpha/2)$ from F_t^* .

4. Remaining Lifetime for Female Medflies in Dependence on Daily Egg-Laying Trajectories

4.1 Estimating Remaining Lifetime

Individual egg-laying counts were recorded daily until death for 1000 mated female medflies (Mediterranean fruit fly, *Ceratitis capitata*) at the medfly mass-rearing facility in Metapa, Mexico. Age-at-death was recorded for each fly. Details of this study are provided in Carey et al. (1998). The aim of the study was to investigate relationships between patterns of reproduction and longevity. One of the basic questions of evolutionary theory is to what extent life extension is driven by enabling increased reproduction. Diverting resources that otherwise would be used for maintenance and repair into reproductive activity may shorten lifespan (Partridge and Harvey, 1985; see also the discussion in Westendorp and Kirkwood, 1999). Data from 893 medflies whose total number of eggs was not less than 20 were entered into this analysis.

We use local linear weighted least squares for smoothing the mean function and the variance–covariance surface as described above. The bandwidth for smoothing the mean function from one-curve-leave-out cross-validation was 2.4d (in days), which we rounded up to take $h = 3d$. The bandwidths for smoothing the covariance function were visually chosen

as $[10d, 10d]$, as cross-validation led to an undersmoothing choice. We define the evolution of the j th eigenfunction over time t as the function family

$$\{[s, t, \hat{\rho}_{jt}(s)], \quad 0 \leq s \leq t, 0 \leq t \leq \tau\}. \quad (37)$$

The time-evolutions of the mean egg-laying function and of the first two eigenfunctions are displayed in Figure 2 (for $t = 30, 40, 50$ days, respectively). These components describe the time-evolution of egg-laying trajectories. We find that mean and eigenfunction evolution is quite smooth, essentially appending new parts smoothly to the previous functions as their domain increases with increasing t , while not dramatically altering function shape on previously included domains. In particular the first eigenfunction evolution is very stable, while the second eigenfunction shows a small amount of flattening for increasing domains. The resulting smooth behavior of the varying coefficients $\hat{\beta}_{jt}$ is confirmed in Figure 3.

While smooth behavior is expected for the mean function evolution, smoothness of the eigenfunction evolution demonstrates that the short-range covariances are relatively stable over time. For all domains, the first eigenfunction delineates a broad rapidly rising peak and subsequent slow decline. The second eigenfunction is positively associated with a sharp and narrower peak with subsequent more rapid decline, and negatively with a less sharp peak followed by extensive egg-laying, a correlation that is biologically highly plausible as the “cost of reproduction,” shown here as a cost of early rapid reproduction. The average leave-one-out RSPEs (30) for QJUE regression for possible candidate values $M = 1, 2, 3, 4, 5$ for

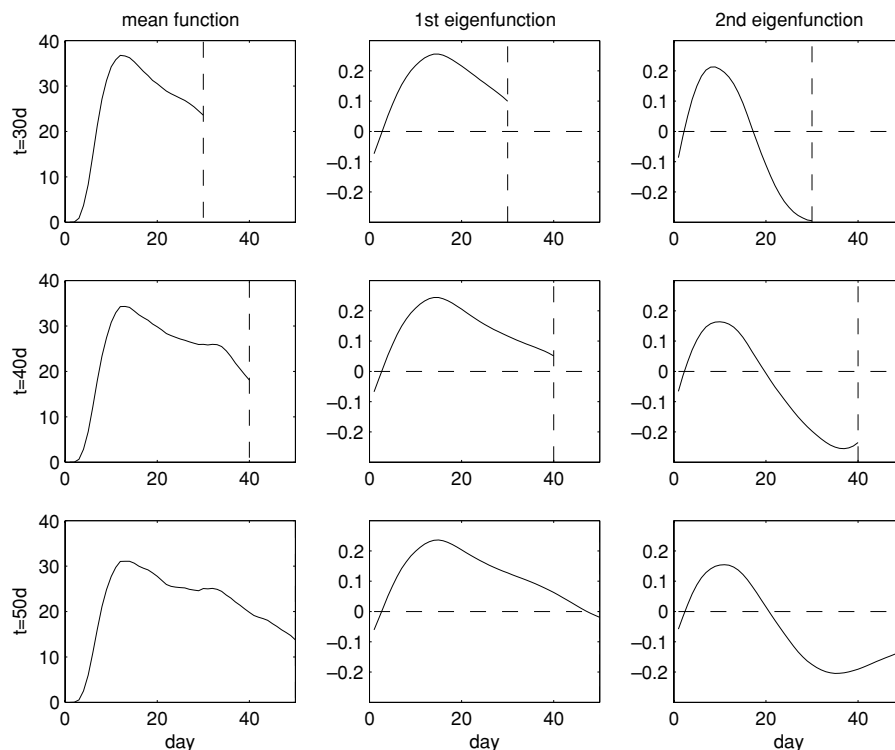


Figure 2. Evolution of mean functions (left column), and of first (middle column) and second (right column) eigenfunctions for current times $t = 30$ (first row), 40 (second row), and 50 (third row) days, for daily egg-laying data.

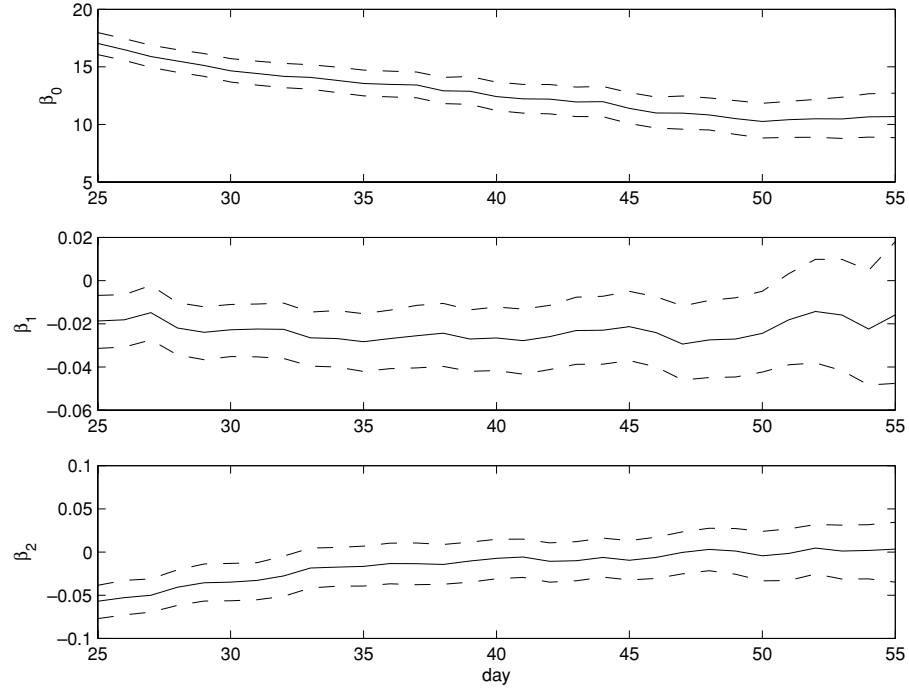


Figure 3. The 95% pointwise bootstrap confidence intervals based on $B = 1000$ bootstrap samples for the varying coefficients β_{0t} (top panel), β_{1t} (middle panel), and β_{2t} (bottom panel) of the functional linear varying coefficient model (16), for the daily egg-laying data. The solid lines are the coefficient estimates $\hat{\beta}_{0t}$, $\hat{\beta}_{1t}$, and $\hat{\beta}_{2t}$ (26).

the number of included components were 10.22, 10.06, 10.12, 10.20, 10.21, respectively, which suggested the choice $M = 2$.

Once the evolution of mean and eigenfunctions has been determined, estimated functional principal component scores $\hat{\varepsilon}_{ijt}$ for each fly, $0 \leq t \leq T$, are obtained via (25). These serve as predictors in various regression models that can be considered for predicting remaining lifetime:

- Cox proportional hazards regression, using current egg-laying values; this is a standard method of survival analysis that has been implemented in many software packages (compare, e.g., Klein and Moeschberger, 1997). At each fixed time point t , we use the current value of the covariate as predictor and fit a Cox proportional hazards model

$$\lambda_i^t(s) = \lambda_0^t(s) \exp\{X_i(t)\gamma_t\}, \quad s \geq t \quad (38)$$

by partial likelihood, where $X_i(t)$ is the realization of the random process X for the i th subject at t . After obtaining the estimated coefficients $\hat{\gamma}_t$ at each time point t , the resulting estimated varying-predictor Cox model is given by $\hat{\lambda}^t(s) = \hat{\lambda}_0^t(s) \exp\{X(t)\hat{\gamma}_t\}$. From the hazard function estimates that this model produces, we then obtain mean remaining lifetime function estimates $\hat{r}(t) = \int_0^\infty \hat{F}^t(s) ds / \hat{F}^t(0)$, where $\hat{F}^t(s) = \exp\{-\int_0^s \hat{\lambda}(u) du\}$.

- The functional linear varying coefficient estimates based on model (16).
- The functional time-varying generalized linear model quasi-likelihood regression estimate (QLUE), based on (14), (15). Link and variance functions for this nonpara-

metric quasi-likelihood are estimated from the data (in our application with bandwidths 17 and 2, respectively, determined according to (32)). The egg-laying trajectories are then summarized by estimated linear predictor functions $\hat{\eta}(t)$ (33).

The functional varying coefficient estimates $\hat{\beta}_{jt}$ (26) obtained for the linear model (16) and also the pointwise bootstrap confidence intervals are illustrated in Figure 3. The coefficients (solid lines) are seen to vary smoothly in t . Coefficients $\hat{\beta}_{0t}$ (solid line in upper panel) are an estimate of the overall mean remaining life function $r_0(t)$ in (16) and taper off accordingly with increasing t . Coefficient estimates $\hat{\beta}_{1t}$ (solid line in middle panel) show quite a bit of fluctuation around a small negative value; the bootstrap confidence bands provide evidence that the underlying values are significantly negative most of the time. The higher the cost in terms of diminished remaining lifetime, the larger is the corresponding estimated predictor $\hat{\varepsilon}_{1t}$, which in turn indicates stronger alignment of an individual's trajectory with the first eigenfunction; it is helpful to take another look at the shape of these eigenfunctions (Figure 2) to see that this corresponds to a broadly elevated level of overall egg-laying. In contrast, coefficient estimates $\hat{\beta}_{2t}$ show a clear upward trend from a significantly negative level with the trend tapering off at higher ages. The peaky egg-laying spike that is reflected in the second eigenfunction with rapid subsequent decline is thus seen to be a portender of short remaining lifetime, and more so at younger ages t ; once a fly with the spiky egg-laying feature has made it to higher age, the prognosis is increasingly less affected by the

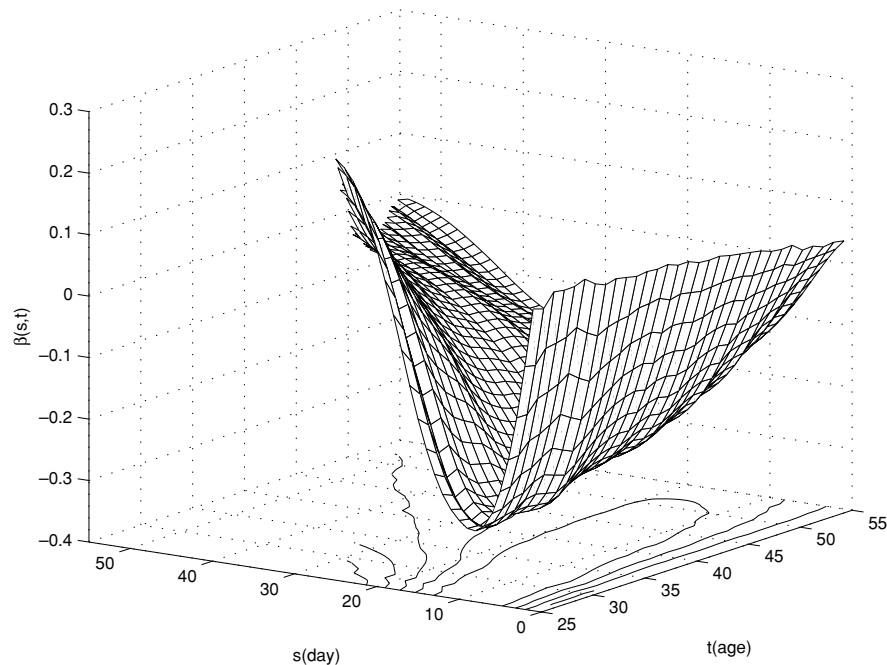


Figure 4. Evolution of the evaluation functions $\beta(s, t)$, $0 \leq s \leq t$, (14) over time in the semiparametric regression model (14) (simultaneous display of surface and contour plots).

spiky peak in early reproduction. Biologically, this could be interpreted as an attenuation of the cost of early reproduction for older flies.

It is also of interest to directly visualize the evolution of the estimate $\hat{\beta}(s, t) = \sum_{j=1}^M \hat{\beta}_{jt} \hat{\rho}_j(s)$ of the time-varying evaluation function $\beta(s, t)$ in (11), i.e., for the generalized linear model (14). This time-evolution is depicted in Figure 4. It is surprisingly smooth. We find that there is a cost of reproduction, as the evaluation function is found to associate reproductive trajectories with higher early peaks with shorter remaining lifespans. Again this effect attenuates for the larger domains and corresponding older ages, as can be seen from the contour plot which indicates a flatter, i.e., less negative, trough for increasing domains. In comparison with the previous analysis of the varying coefficient estimates, this more direct approach may be somewhat easier to interpret, while the conclusions remain similar.

4.2 Comparing Estimates and Estimating Remaining Lifetime Distribution

A baseline in the comparative evaluation of performance of various remaining lifetime predictors is to simply take average remaining lifetime as predictor for individual remaining lifetime; this predictor does not use the covariate process. We present pointwise root squared prediction errors RSPE (30) to compare this baseline method with other predictors for the egg-laying data (Figure 5). Cox regression and the baseline average remaining lifetime clearly have larger and between them quite similar root squared prediction errors than the linear and semiparametric generalized linear varying coefficient estimates; the latter uniformly has the best performance.

Conditional distributions of remaining lifetime, given individual covariate trajectories \tilde{X} up to age t , are of interest

to predict characteristics of remaining lifetime for individuals, for example, by constructing 95% prediction regions. The lower panel of Figure 6 displays the predicted conditional density functions (36) for remaining lifetime at various levels of the linear predictor $\hat{\eta}(t)$ (33) at time $t = 30$ d. Smaller values of $\hat{\eta}(t)$ are seen to be associated with densities that have higher peaks at earlier ages, and therefore, are associated with decreased longevity, while larger values lead to flatter peaks and heavier tails with overall increased longevity. It is interesting to compare this with the observed egg-laying trajectories, sorted according to the value of $\hat{\eta}$, which are displayed in the upper panel of Figure 6. We find that the trajectories with an early rapid rise in egg-laying are those with more peaked densities and reduced survival. The speed of decline in egg-laying also seems to play a role. This adds to the evidence that reproduction is intimately connected with longevity, and that there is a cost in terms of reduced remaining lifetime for early steep increases in egg-laying activity, especially at ages immediately after the reproductive peak.

Revisiting Figure 1, the lower panel displays the estimated conditional density functions of remaining lifetime based on the observed egg-laying trajectories from birth to age 25 (solid segments in upper panel of Figure 1) for two individual flies. In accordance with the overall findings, the egg-laying curve with the higher peak around age 10 leads to a lower longevity prediction and a left-shifted predicted remaining lifetime density. The egg-laying trajectory with the smaller and right-shifted peak of egg-laying activity before age 25 corresponds to increased predicted survival and a density estimate with a much longer right tail and a broader peak. The predicted lifetimes are 33d and 46d, marked as crosses in the figure, and correspond to observed age-at-deaths of 36d and 52d, respectively, marked as open circles.

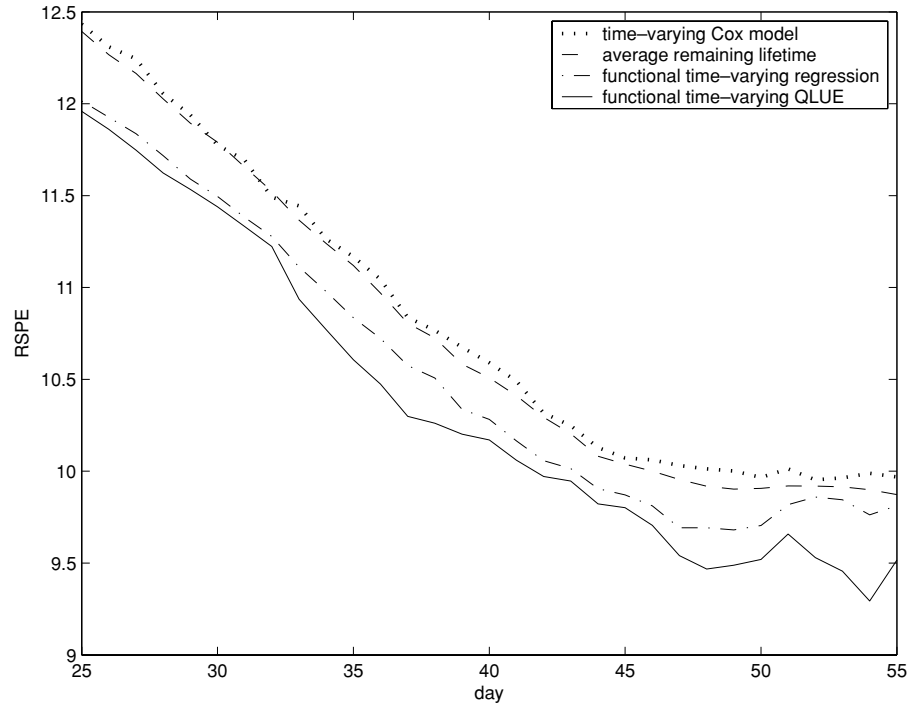


Figure 5. Pointwise average root squared prediction errors (RSPE, (30)) for various predictors of remaining lifetime: time-varying Cox model (38) (dotted), average remaining lifetime (dashed), functional linear varying coefficient model (16) (dash-dot), and semiparametric generalized linear time-varying regression model implemented with the QLUE algorithm (14) (solid), for the daily egg-laying data.

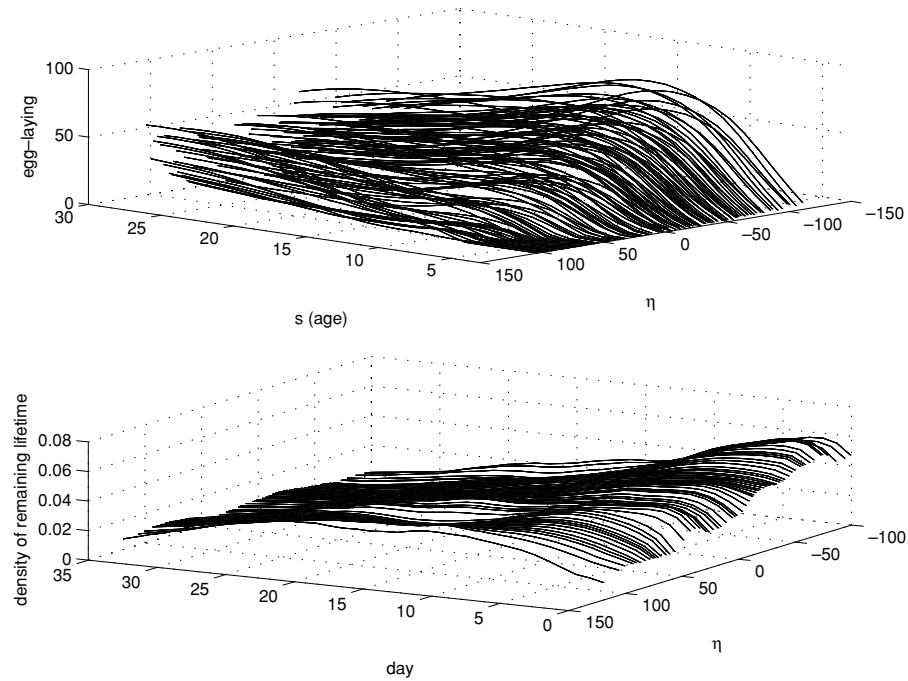


Figure 6. Observed egg-laying curves (upper panel) from birth to current age $t = 30\text{d}$ and estimated densities of remaining lifetime (36) after age 30d (lower panel), both in dependency on the value of the estimated single index function $\hat{\eta}(t)$ (33), obtained from the semiparametric estimating equation (31), for the daily egg-laying data.

5. Discussion and Concluding Remarks

The time-evolution of evaluation functions β (9) and of mean and eigenfunctions is a concept that provides a stepping stone to extend the reach of functional data analysis to the analysis of trajectories that are truncated by death or other events. The interpretation of Figures 2 and 4 depicting this time-evolution and implying an attenuation of the cost of reproduction for older flies indicates that the proposed analysis tools can lead to interesting insights that would be hard to come by with traditional methods. In the medfly example, these insights shed light on the interplay between cost of reproduction and longevity. In other applications it might be a practical matter to predict remaining lifetime and remaining lifetime distributions. In studies on aging, predicted remaining lifetime is a useful measure for senescence (as opposed to age). The proposed methods yield estimates for such measures based on observed covariate trajectories.

In the model comparison reported in Section 4.2, the ubiquitous Cox regression model with time-varying predictors (which takes into account only the last value of the covariate process) is performing relatively poorly and in fact slightly worse than the average remaining lifetime estimator that does not take the covariate into account at all. This demonstrates that the reduction of the information that is contained in the covariate process to just the last observed value is suboptimal in cases where the entire event history plays a role in determining remaining survival. In fact, the proposed methodology indicates that the shape of the egg-laying peak influences cost of reproduction in an age-dependent fashion; such findings would not be possible with methods that only use the current value of the observed trajectories.

Comparing the two better performing models that both take the entire covariate trajectory into account, the semi-parametric generalized linear time-varying approach (11), (31) using the QLUE algorithm, was found to be somewhat better than the linear time-varying model (16), (27). The generalized model allows for more flexibility as it includes smooth link and variance functions, and therefore can be expected to perform better in many situations than the linear model; this increased flexibility turns out to be particularly useful for the estimation of remaining lifetime distributions, where the constraints of dimension reduction through a single index are partly offset by this increased flexibility. At the same time, this approach is more demanding numerically and harder to interpret.

The proposed methods allow for straightforward inclusion of more than one predictor trajectory per subject and also additional multivariate covariates that may be available for each subject in some applications. Additional functional predictors will give rise to additional functional principal scores, and can be easily added to existing predictors in the single index $\eta(t)$ (13), by extending the range of summation. Testing the significance of individual effects may be carried out by bootstrapping; eventually an asymptotic theory might emerge. For the case of functional predictors, this requires not so straightforward increasing dimension asymptotics.

The estimation of conditional distributions of remaining lifetime given a random trajectory as predictor was implemented here under simplifying assumptions, such as assuming a single index as the relevant predictor that can then be

used for nonparametric conditional distribution function estimation with kernels. More complex approaches could be of interest, such as multiple indices. Nearest-neighbor methods hold some promise (Rice, 2004). Once the conditional distribution has been estimated in a satisfactory manner, we may obtain estimates of tail probabilities. The probability that an individual will live beyond a certain age is of actuarial interest and of scientific interest in studies of the mortality of the oldest-old. As we have demonstrated, conditional densities are quite easily obtained and prove useful in demonstrating the effect of covariate trajectories. Other functionals of conditional distributions such as conditional medians and more general conditional quantiles are also easily estimated, leading to notions of conditional quantile regression for functional predictors.

ACKNOWLEDGEMENTS

This research was supported in part by NSF grants DMS02-04869 and DMS03-54448, and NIH grants P01-AG022500-01 and P01-AG08761-10. We are grateful to an associate editor and two referees for their comments which led to many improvements in the article.

REFERENCES

- Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*. New York: Academic Press.
- Bycott, P. and Taylor, J. (1998). A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Statistics in Medicine* **17**, 2061–2077.
- Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association* **95**, 888–902.
- Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics* **30**, 241–255.
- Carey, J., Liedo, P., Müller, H. G., Wang, J. L., and Chiou, J. M. (1998). Relationship of age patterns of fecundity to mortality, longevity and lifetime reproduction in a large cohort of Mediterranean fruit fly females. *Journal of Gerontology, Biological Sciences and Medical Sciences* **53**, 245–251.
- 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. *Journal of the American Statistical Association* **93**, 1376–1387.
- Chiou, J. M. and Müller, H. G. (1999). Nonparametric quasi-likelihood. *Annals of Statistics* **27**, 36–64.
- Cox, D. R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–200.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.
- Hall, P. and Müller, H. G. (2003). Order-preserving nonparametric regression, with applications to conditional distribution and quantile function estimation. *Journal of the American Statistical Association* **98**, 598–608.

- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* **55**, 757–796.
- James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society, Series B* **64**, 411–432.
- Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis Techniques for Censored and Truncated Data*. New York: Springer.
- Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Canadian Journal of Statistics* **31**, 115–128.
- Müller, H. G. (2005). Functional modeling and classification of longitudinal data. *Scandinavian Journal of Statistics* **32**, 223–240.
- Müller, H. G. and Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics* **33**, 774–805.
- Müller, H. G. and Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50**, 61–76.
- Müller, H. G., Carey, J. R., Wu, D., Liedo, P., and Vaupel, J. W. (2001). Reproductive potential predicts longevity of female Mediterranean fruit flies. *Proceedings of the Royal Society B* **268**, 445–450.
- Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy. *Science* **296**, 1029–1031.
- Partridge, L. and Harvey, P. H. (1985). Costs of reproduction. *Nature* **316**, 20–21.
- Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. New York: Springer-Verlag.
- Rice, J. (2004). Functional and longitudinal data analysis: Perspectives on smoothing. *Statistica Sinica* **14**, 631–647.
- Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B* **53**, 233–243.
- Self, S. and Pawitan, Y. (1992). Modeling a marker of disease progression and onset of disease. In *AIDS Epidemiology: Methodological Issues*, N. P. Jewell, K. Dietz, and V. T. Farewell (eds), 231–255. Boston: Birkhäuser.
- Staniswalis, J. G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **93**, 1403–1418.
- Tsiatis, A. A., DeGruttola, V., and Wulfsohn, M.S. (1995). Modeling the relationship of survival to longitudinal data measured with error: Application to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90**, 27–37.
- Vaupel, J. W., Carey, J. R., Christensen, K., et al. (1998). Biodemographic trajectories of longevity. *Science* **280**, 855–860.
- Westendorp, R. G. J. and Kirkwood, T. B. L. (1999). Human longevity at the cost of reproductive success. *Nature* **396**, 743–746.
- Wu, C. O. and Yu, K. F. (2002). Nonparametric varying coefficient models for the analysis of longitudinal data. *International Statistical Review* **70**, 373–393.
- Yao, F., Müller, H. G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A., and Vogel, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics* **59**, 676–685.
- Yu, K. M. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association* **93**, 228–237.

Received July 2004. Revised February 2005.

Accepted February 2005.