BIOMETRICS 70, 794–801
December 2014

DOI: 10.1111/biom.12207

Linear Mixed Function-on-Function Regression Models

Wei Wang

Center for Outcomes Research, Children's Hospital of Philadelphia, Civic Center Blvd, Philadelphia, Pennsylvania 19104, U.S.A.
email: wangw@email.chop.edu

SUMMARY. We develop a linear mixed regression model where both the response and the predictor are functions. Model parameters are estimated by maximizing the log likelihood via the ECME algorithm. The estimated variance parameters or covariance matrices are shown to be positive or positive definite at each iteration. In simulation studies, the approach outperforms in terms of the fitting error and the MSE of estimating the "regression coefficients."

KEY WORDS: ECME algorithm; Functional data analysis; Linear mixed effects models; Principal component analysis.

1. Introduction

Functional data are observed in many research fields and various methods have been proposed, for instance (Cardot, Ferraty, and Sarda, 2003; Ramsay and Silverman, 2005; Reiss and Ogden, 2007; Crainiceanu, Staicu, and Di, 2009; Crambes, Kneip, and Sarda, 2009; Goldsmith et al., 2011; Randolph, Harezlak, and Feng, 2012). Easy to use statistical software packages include "fda" (Ramsay, Hooker, and Graves, 2009; Ramsay et al., 2013) and "refund" (Crainiceanu et al., 2013) of R (R Development Core Team, 2013), R functions at the companion website (www.math.univ-toulouse.fr/staph/npfda) of (Ferraty and Vieu, 2006), and "PACE" package (www.stat.ucdavis.edu/PACE) of MATLAB (The MathWorks Inc., 2013). Among the studies, the regression model where both the response and the predictor are functions is of increasing interest. Methods have been developed to investigate their relationship by Ramsay and Silverman (2005), Yao, Müller, and Wang (2005), Ferraty et al. (2011), Ferraty, Keilegom, and Vieu (2012), and Scheipl, Staicu, and Greven (2013). In this article, we propose a linear mixed modeling approach and study a data set from a biological experiment.

Let $X(\cdot)$ be the predictor function of s with a finite interval domain, and let $Y(\cdot)$ be the response function of t with a finite interval domain. The response $Y(t)$ depends on $X(s)$ through $\int \beta(t,s)X(s)\ \mathrm{d}s$, where the functional coefficient β describes the dependence structure of Y on X. The coefficient $\beta(t,s)$ is assumed smooth and square integrable in the domain, that is, $\int\int \beta^2(t,s)\ \mathrm{d}s\ \mathrm{d}t < \infty$. The model is capable of studying data where the predictor is densely observed and the sample size is relatively small. It extends the scalar-on-function regression model when observations of the response are possibly correlated.

Ferraty et al. (2011, 2012) propose a functional version of the Nadaraya–Watson estimate of the regression operator $\mathrm{E}(Y|X)$. The method is provided in the R function "ffunopare.knn.gcv" where a global bandwidth is selected by the cross validation criterion. In the approach, the estimate of Y is obtained, but $\hat{\beta}$, an estimate of β, is not available. One approach to obtain $\hat{\beta}$ is through estimating the unknown coefficients in the expansion of β by basis functions. Let φ_p's and ψ_q's be two series of basis functions such as splines or eigenfunctions. A tensor product expansion of β follows as

$$\beta(t,s) = \sum_{p=1}^{\infty}\sum_{q=1}^{\infty} b_{pq}\varphi_p(t)\psi_q(s), \tag{1}$$

where the b_{pq}'s are the unknown parameters to be estimated. In practice, the numbers of basis functions are usually truncated as $p \leq P$ and $q \leq Q$.

Ramsay and Silverman (2005) find $\hat{\beta}$ by solving a functional version of the normal equation which is derived from minimizing an integrated residual sum of squares. To use the corresponding R function "linmod" of the "fda" package, one has to specify the smoothing parameter values which are associated with the penalties controlling the smoothness of $\hat{\beta}$. Scheipl et al. (2013) develop a penalized regression model. The smoothing parameters are treated as variance component parameters and estimated by the restricted maximum likelihood method. The proposed method is implemented in the "pffr" function of the "refund" package. Yao et al. (2005) take basis functions as the eigenfunctions and get $\hat{\beta}$ using the eigenvalues and eigenfunctions from smoothed covariance matrices of the X and the Y. The approach has nice asymptotic properties (see also Hall, Müller, and Yao, 2008) and is provided by the "FPCreg" function of the "PACE" package. Peng and Paul (2009) observe that sometimes the estimated variance parameter of the error term in modeling X can be negative.

In this article, we propose a linear mixed effects modeling of the function-on-function regression. Model parameters are estimated by maximizing the log likelihood via the ECME algorithm (Liu and Rubin, 1994). Under very mild conditions, we show at each iteration that the estimated variance parameters are positive and the estimated covariance matrix of the random effects is positive definite. The rest of the article is organized as follows. In Sections 2 and 3, we present the model and the model fitting procedure. We focus on the balanced design case as the data to be analyzed in Section 4 are balanced.

We compare performance of the developed approach (LMFF) with the approaches "ffunopare.knn.gcv" (FF), "pffr" (PFFR) and "FPCreg" (PACE) in data fitting. We also compare with the method proposed by Reiss and Ogden (2007) (FPCR) which is implemented in the R function "fpcr" of "refund." The FPCR approach is proposed in a scalar-on-function regression framework. To apply, we obtain the estimate of Y and β at each t separately. In terms of the mean absolute fitting error, our approach (LMFF) outperforms other methods. In Section 5, we conduct simulations to further compare performance of these approaches. In the settings we tried, the LMFF approach has comparable fitting error of X, smallest fitting error of Y, and most of the times smallest MSE of estimating β. In Section 6, we discuss possible directions for further investigation. Technical derivations are presented in Web Appendices A–C.

2. Model

Using basis function expansions, we present modeling of the response and predictor processes in a linear mixed effects model framework. We then derive the model for the observed data in a vector/matrix form. Unknown "regression coefficients," b_{pq}'s, are contained in the design matrix associated with the random effects. We show the model is identifiable and thus there are no distinct parameters producing the same model.

2.1. *Processes Modeling*

Let ϕ_k's be a series of basis functions. We expand the predictor X_i of the ith individual as $X_i(s) = \mu_x(s) + \sum_{k=1}^{\infty} u_{ik}\phi_k(s)$, where the fixed $\mu_x(s)$ represents the mean and k is truncated at K in practice. In the expansion, we treat the coefficients, u_{ik}'s, random. As we will see in (4), the random coefficients correspond to the random effects in a linear mixed effects model. They can also be viewed as the functional principal component scores as in (Yao et al., 2005) when (i) the ϕ_k's are chosen as the eigenfunctions of the X process, and (ii) u_{ik} and $u_{ik'}$ are uncorrelated if $k' \neq k$. Let x_i be the observed trajectory of X_i contaminated with error ϵ_i. We write

$$x_i(s) = \mu_x(s) + \sum_{k=1}^{K} u_{ik}\phi_k(s) + \epsilon_i(s) \equiv \mu_x(s) + \boldsymbol{\phi}(s)'\mathbf{u}_i + \epsilon_i(s), \tag{2}$$

where $\boldsymbol{\phi}'(s) = (\phi_1(s), \ldots, \phi_K(s))$ and $\mathbf{u}_i = (u_{i1}, \ldots, u_{iK})'$.

Let y_i be the observed value of the response Y_i with mean μ_y. Let y_i depend on X_i through β as $y_i(t) = \mu_y(t) + \int \beta(t,s)\ [X_i(s) - \mathrm{E}X_i(s)]\ \mathrm{d}s + \mathrm{e}_i(t)$, where e_i represents the error. By the expansion of β in (1) and the modeling $X_i = \mu_x(s) + \boldsymbol{\phi}(s)'\mathbf{u}_i$, we then write

$$y_i(t) = \mu_y(t) + \sum_{p=1}^{P}\sum_{q=1}^{Q} b_{pq}\varphi_p(t)\left(\int \psi_q(s)\boldsymbol{\phi}'(s)\mathrm{d}s\right)\mathbf{u}_i + \mathrm{e}_i(t). \tag{3}$$

2.2. *Observed Data*

Suppose the sampling points of the ith individual are s_{ij}, $j = 1, \ldots, n_i$, and $t_{il}, l = 1, \ldots, m_i$. The observed data are then $\mathbf{x}_i = (x_i(s_{i1}), \ldots, x_i(s_{in_i}))'$ and $\mathbf{y}_i = (y_i(t_{i1}), \ldots, y_i(t_{im_i}))'$. Data collected from the N individuals are thus $\{(\mathbf{x}_i, \mathbf{y}_i),\ i = 1, \ldots, N\}$.

Let $\boldsymbol{\mu}_{x,i} = (\mu_x(s_{i1}), \ldots, \mu_x(s_{in_i}))'$ and $\boldsymbol{\epsilon}_i = (\epsilon_i(s_{i1}), \ldots, \epsilon_i(s_{in_i}))'$. From (2), we get

$$\begin{gathered}\mathbf{x}_i = \boldsymbol{\mu}_{x,i} + \mathbf{A}_i\mathbf{u}_i + \boldsymbol{\epsilon}_i,\\ \mathbf{A}_i[j,k] = \phi_k(s_{ij}),\ \mathbf{u}_i \sim (\mathbf{0}, \boldsymbol{\Sigma}_u),\ \boldsymbol{\epsilon}_i \sim (\mathbf{0}, \sigma_\epsilon^2\mathbf{E}_{x,i}).\end{gathered} \tag{4}$$

Model of $\mathbf{x}_i$ in (4) has a linear mixed model representation with fixed effects $\boldsymbol{\mu}_{x,i}$, known design matrix $\mathbf{A}_i$, random effects $\mathbf{u}_i$ with covariance matrix $\boldsymbol{\Sigma}_u$, and error $\boldsymbol{\epsilon}_i$. We assume the error $\boldsymbol{\epsilon}_i$ has a covariance matrix parametrized by σ_ϵ^2 where $\mathbf{E}_{x,i}$ is a known symmetric and positive definite matrix.

Let $\boldsymbol{\mu}_{y,i} = (\mu_y(s_{i1}), \ldots, \mu_y(s_{in_i}))'$ and $\mathbf{e}_i = (\mathrm{e}_i(t_{i1}), \ldots, \mathrm{e}_i(t_{im_i}))'$. From (3), we have

$$\begin{gathered}\mathbf{y}_i = \boldsymbol{\mu}_{y,i} + \boldsymbol{\Phi}_i\mathbf{B}\mathbf{T}\mathbf{u}_i + \mathbf{e}_i,\\ \boldsymbol{\Phi}_i[l,p] = \varphi_p(t_{il}),\ \mathbf{B}[p,q] = b_{pq},\\ \mathbf{T}[q,k] = \int \psi_q(s)\ \phi_k(s)\ \mathrm{d}s,\ \mathbf{e}_i \sim (\mathbf{0}, \sigma_e^2\mathbf{E}_{y,i}),\end{gathered} \tag{5}$$

where the covariance matrix of $\mathbf{e}_i$ is parametrized by σ_e^2 and $\mathbf{E}_{y,i}$ is also known, symmetric and positive definite. Similarly, $\mathbf{y}_i$ in (5) has a linear mixed model representation but with a partially known design matrix $\boldsymbol{\Phi}_i\mathbf{B}\mathbf{T}$ where the unknown $\mathbf{B}$ is to be estimated.

Let $\mathbf{z}_i' = (\mathbf{x}_i', \mathbf{y}_i')$. We obtain the joint model of $\mathbf{x}_i$ and $\mathbf{y}_i$ as

$$\begin{gathered}\mathbf{z}_i = \begin{pmatrix}\mathbf{x}_i\\ \mathbf{y}_i\end{pmatrix} = \boldsymbol{\mu}_{z,i} + \mathbf{D}_i\mathbf{u}_i + \boldsymbol{\epsilon}_{e_i},\\ \boldsymbol{\mu}_{z,i} = \begin{pmatrix}\boldsymbol{\mu}_{x,i}\\ \boldsymbol{\mu}_{y,i}\end{pmatrix},\ \mathbf{D}_i = \begin{pmatrix}\mathbf{A}_i\\ \boldsymbol{\Phi}_i\mathbf{B}\mathbf{T}\end{pmatrix},\ \boldsymbol{\epsilon}_{e_i} = \begin{pmatrix}\boldsymbol{\epsilon}_i\\ \mathbf{e}_i\end{pmatrix},\\ \mathbf{u}_i \sim (\mathbf{0}, \boldsymbol{\Sigma}_u),\quad \boldsymbol{\epsilon}_{e_i} \sim (\mathbf{0}, \mathbf{E}_i),\quad \mathbf{E}_i = \begin{pmatrix}\sigma_\epsilon^2\mathbf{E}_{x,i} & \mathbf{0}\\ \mathbf{0} & \sigma_e^2\mathbf{E}_{y,i}\end{pmatrix}.\end{gathered} \tag{6}$$

The model of $\mathbf{z}_i$ has unknown parameters $\{\boldsymbol{\mu}_{z,i}, \boldsymbol{\Sigma}_u, \sigma_\epsilon^2, \sigma_e^2, \mathbf{B}\}$. In the following, we take $\boldsymbol{\Sigma}_u$ as an unstructured covariance matrix. In the data analysis and simulations, we take $\mathbf{E}_{x,i} = \mathbf{I}_{n_i}$ and $\mathbf{E}_{y,i} = \mathbf{I}_{m_i}$. We discuss the selection of other covariance matrix structures in Section 6. It follows that

$$\begin{aligned}\mathrm{Cov}(\mathbf{z}_i) \equiv \boldsymbol{\Sigma}_{z_i} &= \mathbf{D}_i\boldsymbol{\Sigma}_u\mathbf{D}_i' + \mathbf{E}_i\\ &= \begin{pmatrix}\mathbf{A}_i\boldsymbol{\Sigma}_u\mathbf{A}_i' + \sigma_\epsilon^2\mathbf{E}_{x,i} & \mathbf{A}_i\boldsymbol{\Sigma}_u\mathbf{T}'\mathbf{B}'\boldsymbol{\Phi}_i'\\ \boldsymbol{\Phi}_i\mathbf{B}\mathbf{T}\boldsymbol{\Sigma}_u\mathbf{A}_i' & \boldsymbol{\Phi}_i\mathbf{B}\mathbf{T}\boldsymbol{\Sigma}_u\mathbf{T}'\mathbf{B}'\boldsymbol{\Phi}_i' + \sigma_e^2\mathbf{E}_{n,i}\end{pmatrix}.\end{aligned}$$

We assume the random parts in the model are all mutually independent. As in standard linear mixed models, we also assume they are all normally distributed; for instance (Laird and Ware, 1982). Unknown parameters in the joint model of all individuals' are then $\boldsymbol{\theta} = \{\boldsymbol{\mu}_{z,1}, \ldots, \boldsymbol{\mu}_{z,N}, \boldsymbol{\Sigma}_u, \sigma_\epsilon^2, \sigma_e^2, \mathbf{B}\}$.

2.3. *Identifiability*

A multivariate normal distribution is uniquely characterized by its mean vector and covariance matrix. Since all the $\mathbf{z}_i$'s are normally distributed, we only need to check if the parameters in the mean vectors and covariance matrices are identifiable. It is easy to see there is no $\boldsymbol{\mu}_{z,i}^* \neq \boldsymbol{\mu}_{z,i}$ giving the same mean vector of $\mathbf{z}_i$. As all the $\mathbf{z}_i$'s have the same covariance parameters $\{\boldsymbol{\Sigma}_u, \sigma_\epsilon^2, \sigma_e^2, \mathbf{B}\}$, we need to check if they are identifiable in an arbitrary $\mathbf{z}_i$'s model (Wang, 2013). We show in Web Appendix A that these covariance parameters are identifiable under the assumptions that $\mathbf{T}$ is invertible, and the $\mathbf{A}_i$ and the $\boldsymbol{\Phi}_i$ are of full column rank. The conditions are easy to be satisfied as $\mathbf{T}$ is a low dimensional matrix with elements being the integration of basis functions, and usually we have $n_i \gg K$ and $m_i \gg P$.

3. Model Fitting

We present the model fitting procedure in the balanced design case. Extension to fit an unbalanced design is discussed in Web Appendix C. We have $n_i \equiv n$, $s_{ij} = s_j$, $m_i \equiv m$, $t_{il} = t_l$. It follows that $\{\boldsymbol{\mu}_{z,i}, \mathbf{D}_i, \boldsymbol{\Sigma}_{z_i}\}$ are the same for all i. To save notation, we suppress the subscript i in the following. To fit the model, we first choose the basis functions ϕ_k's, φ_p's and ψ_q's contained in the design matrix $\mathbf{D}$. In this article, we use the basis functions as the eigenfunctions estimated from the sample covariance matrices. Specifically, we use $\phi_1, \ldots, \phi_K$ in (2) equal to the first K estimated eigenfunctions from the sample covariance matrix of the $\mathbf{x}_i$'s. As in (Yao et al., 2005), we use $\varphi_1, \ldots, \varphi_P$ in (1) equal to the first P estimated eigenfunctions of the $\mathbf{y}_i$'s, and take $\psi_j = \phi_j$ and $Q = K$. We then treat the basis functions known and estimate the model parameters $\boldsymbol{\theta} = \{\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_u, \sigma_\epsilon^2, \sigma_e^2, \mathbf{B}\}$ by maximizing the log-likelihood of the observed data $\{\mathbf{z}_i \equiv (\mathbf{x}_i, \mathbf{y}_i),\ i = 1, \ldots, N\}$. To assess the model fit, we look at the fitted values of $\mathbf{x}_i$ and $\mathbf{y}_i$ based on the BLUP of $\mathbf{u}_i$.

3.1. *Parameter Estimates*

Closed form of the estimate of $\boldsymbol{\mu}_z$ can be obtained by maximizing the log-likelihood directly. From (6), it is clear that the MLE of $\boldsymbol{\mu}_z$ is $\hat{\boldsymbol{\mu}}_z = \bar{\mathbf{z}}$. We use the EM type algorithm (Laird and Ware, 1982; Meng and Rubin, 1993) to estimate $\{\boldsymbol{\Sigma}_u, \sigma_\epsilon^2, \sigma_e^2, \mathbf{B}\}$ and the overall estimation procedure is in the spirit of the ECME algorithm (Liu and Rubin, 1994). To use the algorithm, we treat $\{(\mathbf{z}_i, \mathbf{u}_i),\ i = 1, \ldots, N\}$ as the complete data. In the E-step, given the previous estimate $\hat{\boldsymbol{\theta}}$, we calculate the conditional expectation of the log-likelihood of the complete data on the observed data. In the M-step, we maximize the conditional log-likelihood with respect to each of $\{\boldsymbol{\Sigma}_u, \sigma_\epsilon^2, \sigma_e^2, \mathbf{B}\}$ given the others fixed at their previous or current estimates.

The E-step calculation involves the following two terms

$$\begin{aligned}\boldsymbol{\mu}_{u_i|z_i} &\equiv \mathrm{E}(\mathbf{u}_i|\mathbf{z}_i) = \boldsymbol{\Sigma}_u \mathbf{D}' \boldsymbol{\Sigma}_z^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_z) \\ \boldsymbol{\Sigma}_{u_i|z_i} &\equiv \mathrm{Cov}(\mathbf{u}_i|\mathbf{z}_i) = \boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_u \mathbf{D}' \boldsymbol{\Sigma}_z^{-1} \mathbf{D} \boldsymbol{\Sigma}_u, \end{aligned} \tag{7}$$

evaluated at the previous estimate $\hat{\boldsymbol{\theta}}$. Expressions of the current parameter estimate are given in the following and details of the derivation are presented in Web Appendix B. In Web Appendix B, we also justify that the estimate of $\boldsymbol{\Sigma}_u$ is positive definite and the estimates of $\{\sigma_\epsilon^2, \sigma_e^2\}$ are positive at each iteration if the starting values are.

The estimate of $\boldsymbol{\Sigma}_u$ is

$$\boldsymbol{\Sigma}_u = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{\mu}}_{u_i|z_i} \hat{\boldsymbol{\mu}}_{u_i|z_i}' + \hat{\boldsymbol{\Sigma}}_{u|z}. \tag{8}$$

Let

$$\hat{s}_{x,i} = (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_x - \mathbf{A}\hat{\boldsymbol{\mu}}_{u_i|z_i})' \mathbf{E}_x^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_x - \mathbf{A}\hat{\boldsymbol{\mu}}_{u_i|z_i}). \tag{9}$$

We obtain

$$\sigma_\epsilon^2 = \frac{1}{nN} \sum_{i=1}^{N} \hat{s}_{x,i} + \frac{1}{n} \mathrm{tr}[\mathbf{E}_x^{-1} \mathbf{A} \hat{\boldsymbol{\Sigma}}_{u|z} \mathbf{A}']. \tag{10}$$

Let "vec" be the operator which transforms a matrix into a vector by stacking the columns one underneath another. Let $\otimes$ be the Kronecker product which maps two arbitrarily dimensioned matrices into a larger matrix with special block structure. We have

$$\begin{aligned}\mathrm{vec}(\mathbf{B}) = {} & \left[\left(\mathbf{T} \sum_{i=1}^{N} \hat{\boldsymbol{\mu}}_{u_i|z_i} \hat{\boldsymbol{\mu}}_{u_i|z_i}' \mathbf{T}' + N \mathbf{T} \hat{\boldsymbol{\Sigma}}_{u|z} \mathbf{T}' \right) \otimes \left(\boldsymbol{\Phi}' \mathbf{E}_y^{-1} \boldsymbol{\Phi} \right) \right]^{-1} \\ & \times \mathrm{vec} \left(\sum_{i=1}^{N} \boldsymbol{\Phi}' \mathbf{E}_y^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_y) \hat{\boldsymbol{\mu}}_{u_i|z_i}' \mathbf{T}' \right).\end{aligned}$$

From (1), the estimate of β is then

$$\beta(t, s) = \sum_{p=1}^{P} \sum_{q=1}^{Q} b_{pq} \varphi_p(t) \psi_q(s) \equiv \boldsymbol{\varphi}'(t) \mathbf{B} \boldsymbol{\psi}(s).$$

Let $\hat{s}_{y,i}(\mathbf{B}) = (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_y - \boldsymbol{\Phi}\mathbf{B}\mathbf{T}\hat{\boldsymbol{\mu}}_{u_i|z_i})' \mathbf{E}_y^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_y - \boldsymbol{\Phi}\mathbf{B}\mathbf{T}\hat{\boldsymbol{\mu}}_{u_i|z_i})$. We get

$$\sigma_e^2 = \frac{1}{mN} \sum_{i=1}^{N} \hat{s}_{y,i}(\mathbf{B}) + \frac{1}{m} \mathrm{tr}[\mathbf{E}_y^{-1} \boldsymbol{\Phi} \mathbf{B} \mathbf{T} \hat{\boldsymbol{\Sigma}}_{u|z} \mathbf{T}' \mathbf{B}' \boldsymbol{\Phi}'].$$

4. Data Analysis

We apply the LMFF approach to analyze a data set from an evolutionary biology experiment where mice were bred over generations. Researches have identified a negative age-specific correlation between the wheel running distance and the mouse body mass. We are interested to study the dependence of the wheel running distance, $\mathbf{y}_i$'s, on the log body mass, $\mathbf{x}_i$'s, simultaneously for all ages of the population of $N = 38$ mice. Figure 2 contains the plot of $\hat{\beta}$ and Table 1 shows the fitting errors for different approaches. Compared with other methods, the LMFF approach has smaller errors.

Table 1
Comparison of the mean absolute fitting errors of log body mass (x) and of wheel running distance (y)

		Fitting error (x)			Fitting error (y)				
K	P	LMFF	PACE	PFFR	LMFF	PACE	PFFR	FF	FPCR
2	5	0.0223	0.0301	0.0435	12.1925	12.5731	12.5449	13.2169	12.5520
5	5	0.0168	0.0286	0.0412	7.7596	11.5153	11.4097	13.1225	11.6656

4.1. *Experiment*

Selective-breeding experiments were performed on house mice (Mus domesticus) for generations for increased voluntary wheel-running exercise (Swallow, Carter, and Garland, 1998; Garland, 2003). Mice were placed individually in cages with a 11.75 cm radius running wheel and electronic wheel-revolution counter built into the cage-top. The mouse had the option of voluntarily getting into the wheel and running, or of remaining in the cage and not running. Once a week each mouse was weighed and its weekly wheel revolutions were downloaded from the counter device. Cages were cleaned weekly and running wheels were cleaned monthly. Monthly wheel running distance was obtained based on weekly wheel revolutions.

Mice were selected for high voluntary wheel running activity at 8 weeks of age after 10 generations of selection (Swallow et al., 1998). A subsequent study, 14 generations after selection, examined wheel running and body mass in this system of mice after 8 weeks of access to running wheels (Swallow et al., 1999). The study at the younger ages (weeks 8–15) finds a negative genetic correlation between age-specific body mass and wheel running, and access to running wheels did result in a decrease in body mass. Morgan, Garland, and Carter (2003) conducted a detailed study on body mass and wheel running ontogenies in the 16th generation of this system of mice. The authors identified that selected mice showed a steeper decline in wheel running activity at older ages, age-specific body mass was a significant predictor for age-specific wheel running, and smaller individuals ran more than larger individuals.

4.2. *Estimate of β*

The log body mass trajectories of $n = 58$ weeks are shown in the upper left panel of Figure 1. Overall, the body mass is an increasing function of age with a sharp rate at early weeks (<10). The upper right panel shows the wheel running distances of $m = 15$ months for each mouse. Compared with the log body mass, the wheel running distances are irregular and variable across individuals. The corresponding lower two panels show the first two eigenfunctions obtained from sample covariance matrices of log body mass and wheel running separately. The first eigenfunctions of log body mass and wheel running distance are both close to constants indicating that the biggest source of variability is the overall size of the mouse or the overall wheel running activity. The second eigenfunction contrasts body mass before 5 weeks or wheel running before 4 months to later weight or wheel running.

We choose the number of eigenfunctions K and P by looking at the proportion of cumulative variance (90%). In the log body mass, the first two eigenfunctions explain 91% of the cumulative variance ($K = 2$). In the wheel running distance, the first five eigenfunctions explain 90% of the cumulative variance ($P = 5$). Figure 2 shows the $\hat{\beta}(s, t)$ at each week (s) and month (t). We observe a high dependence of wheel running activity on early body mass (<10 weeks) uniformly across all months. The dependence decays and becomes negative in later weeks. The result of older ages is consistent with those of Swallow et al. (1999) and Morgan et al. (2003) who measured a negative age-specific relationship between wheel running and body mass. Our study also reveals a positive association between wheel running activity and early age body mass. In the experiment, the mice were selected for high wheel running activity at 8 weeks for generations. The result suggests that young larger mice were more active than young smaller mice.

Figures of the residuals from the model fit are presented in Web Figure 1. The plots show a random pattern of homogeneity versus time and the fitted values. We also fit the model with $K = 2, 3, 4$, and $P = 2, 5$, and observe similar pattern of $\hat{\beta}(s, t)$ in all of the six combinations (Web Figure 2).

4.3. *Confidence Interval*

We build a point wise 95% confidence interval at each (t, s) of β as $\hat{\beta}(t, s)$ plus or minus its standard deviation multiplied by 1.96. To find the standard deviation, we use the bootstrap method with 1000 bootstrap samples by resampling the $\mathbf{z}_i$'s. It turns out only 18 of the 15×58 confidence intervals do not contain 0. In this generation of mice, we observe a consistent pattern of the dependence of wheel running activity on the log body mass. However, only 2% of the point wise patterns is significant.

To evaluate the coverage of the constructed confidence interval, we simulate data of the same sample size 38 from the fitted model and run 1000 simulations to calculate the proportion of covering $\hat{\beta}$ at each (t, s). We obtain a mean proportion of coverage as 0.940 with a standard deviation of 0.011. Quartiles of the coverage proportion are 0.899 (0%), 0.933 (25%), 0.941 (50%), 0.948 (75%), 0.968 (100%). Contour plot of the coverage proportion at each (t, s) and the overall histogram is provided in Web Figure 3.

4.4. *Fitting Error*

We assess the model fit by looking at the mean absolute fitting error of the $\mathbf{x}_i$'s and the $\mathbf{y}_i$'s. We first compute the averaged absolute errors over the sampling points and then look at the mean of the averaged errors across all individuals. We also choose $K = 5$ and $P = 5$ by the AIC_R and BIC criteria of the PACE method. We apply the PACE, PFFR, FF, and FPCR methods to fit the data. Unlike LMFF or PACE, the PFFR, FF and FPCR approaches are not using information from

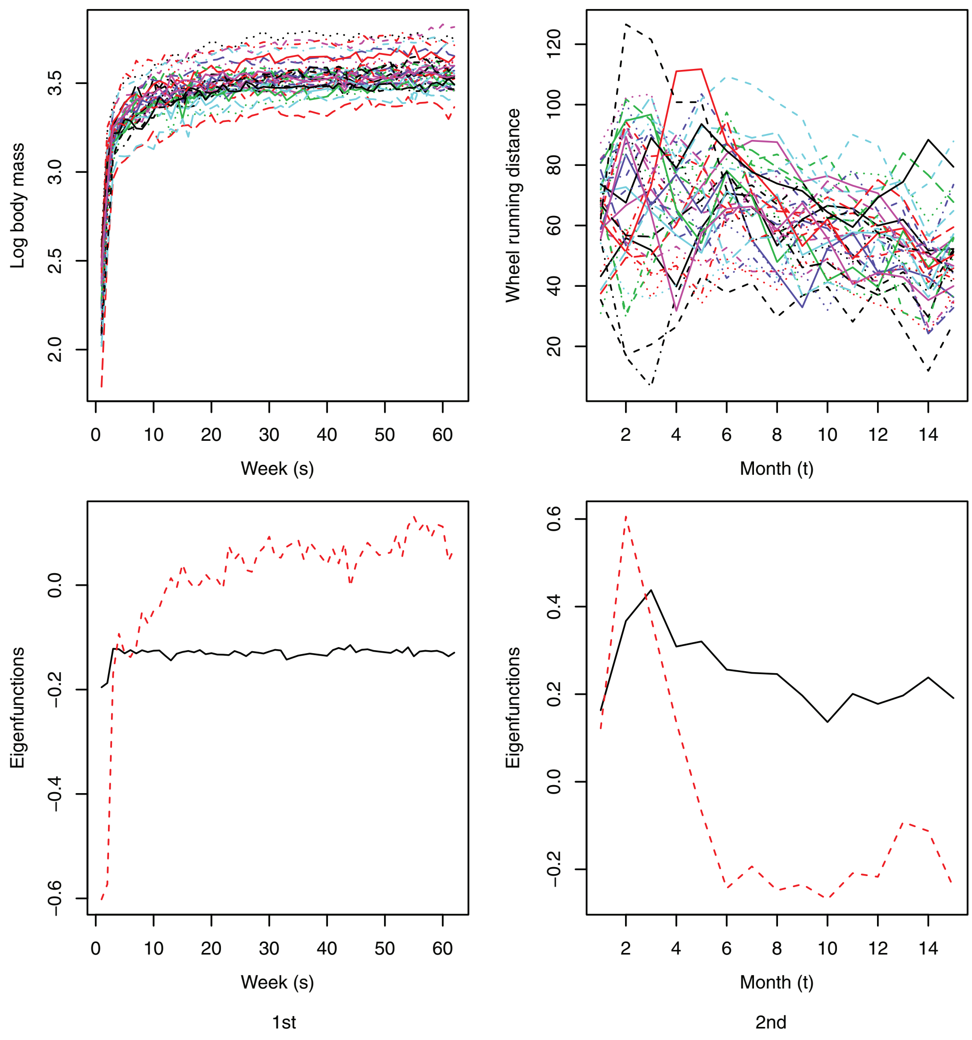

Figure 1. Log body mass of 58 weeks and wheel running distances of 15 months (each line corresponds to one mouse), and their corresponding first two eigenfunctions.

the eigenvalue decomposition of the covariance matrix of the $\mathbf{y}_i$'s. In the application, we adopt the default settings of the corresponding R or MATLAB functions. Table 1 summarizes the fitting errors of different approaches, where we see LMFF has the smallest fitting errors.

5. Simulations

We conduct simulations to compare performance of the various approaches. The random effects are simulated with a diagonal covariance matrix structure which is assumed by the PACE method. We fit an unstructured $\boldsymbol{\Sigma}_u$ in the implementation of LMFF. We run 1000 simulations for sample sizes $N = 100$ and $N = 200$. For ease, in each simulation, K and P are chosen by the PACE AIC$_R$ and BIC criteria respectively. Two thirds and one-third of the chosen K and P are also used.

We take s as consecutive integers from 1 to 30 divided by 15, and take t as consecutive integers from 1 to 15 divided by 10. We set $\mu_x(s) = s + \sin(s)$ and $\mu_y(t) = 1.5\ln(t+1)$. To simulate x_i, in (2), we let $\phi_1(s) = -\cos(\pi s/2)/\sqrt{5}$ and $\phi_2(s) = \sin(\pi s/2)/\sqrt{5}$. We use two β functions: $\beta_1(t, s) = s\cos(\pi|s - t|)/3$ and $\beta_2(t, s) = [\cos(\pi s)\sin(\pi t) + (st)^2]/13$. Plots of β_1 and β_2 are provided in Web Figure 4.

Variance parameter values are $\sigma_\epsilon^2 = 0.2$ and $\sigma_e^2 = 0.1$. We let the covariance matrices of the errors $\mathbf{E}_x$ and $\mathbf{E}_y$ both be identity matrices, and both be correlated as $\mathbf{E}_x(s, s^*) = e^{-5\sqrt{|s-s^*|}}$, $\mathbf{E}_y(t, t^*) = 0.1^{5|t-t^*|}$. We simulate $\mathbf{u}_i$ from a normal distribution, and from a mixture of two normal distributions. The normal distribution has mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_u = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$. The two normal distributions have mean vectors $(1\ 1/\sqrt{2})'$ and $-(1\ 1/\sqrt{2})'$, and the same covariance matrix $\boldsymbol{\Sigma}_u/2$. The random effects $\mathbf{u}_i$ can come from either distribution with probability 1/2.

In each simulation, we first compute the averaged absolute errors of fitting the x_i and the y_i over s or t. Then we look at the mean of the averaged errors across the individuals. We also compute the squared error of estimating β at each (t, s) and then obtain the sum of the squared error divided by 30×15. As a summary, we get the mean absolute fitting error and mean squared error (MSE) of estimating β over the

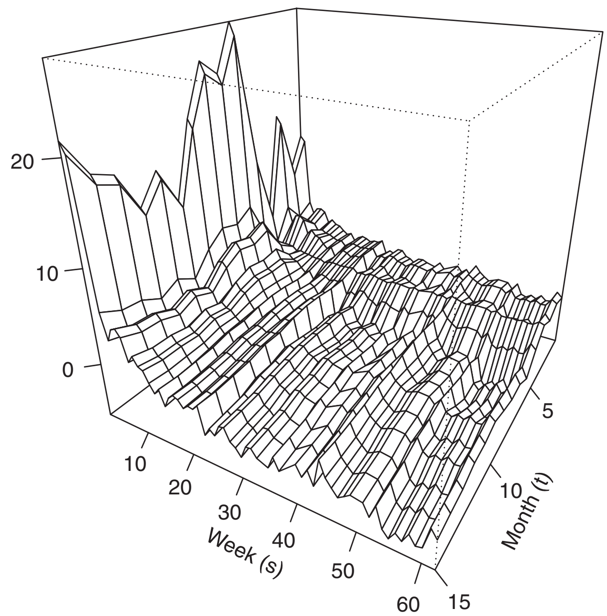

Figure 2. Estimate of $\beta(s,t)$ at each week s and month t.

1000 simulations. The results of normally distributed random effects are summarized in Table 2 for identity $\mathbf{E}_x$ and $\mathbf{E}_y$, and in Table 3 for correlated $\mathbf{E}_x$ and $\mathbf{E}_y$. We observe very similar results when the random effects follow a mixture distribution (Web Tables 1 and 2). In the tables, the K and P columns show the averaged number of eigenfunctions used over the 1000 simulations. In each simulation, the numbers are chosen by the PACE method, and then two thirds and one-third of the numbers are also used.

In general, using more eigenfunctions reduces the fitting errors for all approaches. In the comparison, the smallest fitting error or MSE is in bold font. The LMFF and PACE approaches always outperform except that the FPCR approach has the smallest MSE of estimating β_2 when $N = 100$ and all chosen eigenfunctions are used. The LMFF approach performs best in fitting the response and has comparable fitting error of the predictor. In these settings, both approaches have the smallest MSE when a moderate number of the chosen eigenfunctions (two thirds) is used. Over-fitting or under-fitting is observed when more or less eigenfunctions are used.

In summary, the LMFF approach always has smallest fitting error of the response, and smaller fitting error of the predictor in about half of the settings compared to the PACE approach. The LMFF approach also has the smallest MSE of estimating β in the majority of the settings.

6. Discussion

We develop a linear mixed regression model where both the response and predictor are functions (LMFF). In simulations, the LMFF approach has smallest fitting error of the response, and most of the times smallest MSE of estimating the "regression coefficient" β. The fitting error of the predictor is comparable to that of the PACE method. There are some similarities between LMFF and PACE. In PACE, the response and the predictor are represented through Karhunen–Loéve expansions. In the expansion, the random coefficients correspond to the random effects in linear mixed modeling with a diagonal covariance matrix structure. Normality assumption is invoked to find the fitted values of the response in PACE.

The major distinction between the two approaches is that LMFF is parametric likelihood based. In PACE, one first gets estimated values separately from the smoothed covariance matrices of the predictor, of the response and between the predictor and the response. Based on these estimated values, $\hat{\beta}$ is then obtained. In LMFF, we write a joint log likelihood for the predictor and the response, and estimate the model parameters through maximizing the log likelihood. We use the ECME algorithm for maximization and under regularity conditions the ECME estimate converges to the maximum likelihood estimate (Liu and Rubin, 1994). Properties of $\hat{\beta}$ then inherit from the MLE properties.

We assume the random effects in the model follow a normal distribution to obtain the corresponding log likelihood. In standard linear mixed models, studies have shown that the

Table 2

Comparison of the fitting errors and the MSE of estimating β. The random effects are simulated from a normal distribution and the error covariance structures are both identity.

				Fitting error (x)			Fitting error (y)					MSE (β)			
β	N	K	P	LMFF	PACE	PFFR	LMFF	PACE	PFFR	FPCR	FF	LMFF	PACE	PFFR	FPCR
β_1	100	2.381	2.725	0.3372	**0.3032**	0.3440	**0.2485**	0.2548	0.2509	0.2501	0.2611	0.0634	**0.0628**	0.3645	0.0751
		2.155	2.124	**0.3395**	0.3422	0.3445	**0.2495**	0.2553	0.2510	0.2502	0.2611	0.0527	**0.0521**	0.2779	0.0751
		1.107	1.143	**0.4168**	0.4175	0.4187	**0.2536**	0.2589	0.2557	0.2550	0.2635	**0.0538**	0.0539	0.1837	0.0745
	200	2.280	2.816	0.3412	**0.3176**	0.3443	**0.2503**	0.2562	0.2523	0.2518	0.2590	**0.0536**	0.0598	0.2805	0.0751
		2.100	2.168	**0.3426**	0.3439	0.3446	**0.2508**	0.2564	0.2524	0.2519	0.2590	**0.0476**	0.0498	0.2562	0.0751
		1.069	1.157	**0.4224**	**0.4224**	0.4230	**0.2552**	0.2598	0.2571	0.2568	0.2620	**0.0526**	0.0535	0.1825	0.0745
β_2	100	2.502	2.239	0.3356	**0.3032**	0.3437	**0.2483**	0.2531	0.2507	0.2504	0.2586	0.0272	0.1015	0.2142	**0.0234**
		2.227	1.876	**0.3384**	0.3417	0.3442	**0.2489**	0.2534	0.2507	0.2504	0.2586	0.0157	**0.0139**	0.1016	0.0234
		1.140	1.071	**0.4146**	0.4157	0.4169	**0.2533**	0.2569	0.2548	0.2542	0.2608	**0.0155**	0.0161	0.0438	0.0243
	200	2.385	2.268	0.3402	**0.3176**	0.3441	**0.2504**	0.2544	0.2520	0.2519	0.2571	**0.0184**	0.0223	0.1143	0.0233
		2.165	1.893	**0.3418**	0.3435	0.3445	**0.2507**	0.2545	0.2520	0.2519	0.2571	**0.0109**	0.0115	0.0652	0.0233
		1.108	1.057	**0.4192**	0.4195	0.4201	**0.2549**	0.2580	0.2562	0.2559	0.2598	**0.0147**	0.0160	0.0431	0.0243

Table 3

Comparison of the fitting errors and the MSE of estimating β. The random effects are simulated from a normal distribution and the error covariance structures are correlated.

				Fitting error (x)			Fitting error (y)					MSE (β)			
β	N	K	P	LMFF	PACE	PFFR	LMFF	PACE	PFFR	FPCR	FF	LMFF	PACE	PFFR	FPCR
β_1	100	2.089	6.251	0.2855	**0.1846**	0.2873	**0.2503**	0.2543	0.2536	0.2534	0.2638	0.0541	**0.0478**	0.2024	0.0747
		1.972	4.340	**0.2878**	0.2890	0.2896	**0.2512**	0.2551	0.2538	0.2535	0.2639	**0.0469**	**0.0469**	0.1944	0.0748
		1.024	2.334	0.4046	**0.4039**	0.4045	**0.2549**	0.2597	0.2581	0.2577	0.2657	**0.0531**	0.0538	0.1564	0.0743
	200	2.179	6.941	0.2804	**0.1912**	0.2810	**0.2523**	0.2556	0.2549	0.2547	0.2617	0.0504	**0.0456**	0.2012	0.0748
		2.029	4.942	**0.2833**	0.2838	0.2839	**0.2528**	0.2561	0.2551	0.2547	0.2617	**0.0446**	0.0453	0.1961	0.0749
		1.022	2.836	0.4090	**0.4079**	0.4082	**0.2556**	0.2600	0.2589	0.2587	0.2636	**0.0515**	0.0524	0.1598	0.0744
β_2	100	1.886	6.252	0.3072	**0.1846**	0.3090	**0.2501**	0.2530	0.2526	0.2526	0.2603	0.0177	**0.0116**	0.0380	0.0238
		1.799	4.342	**0.3089**	0.3100	0.3106	**0.2509**	0.2535	0.2527	0.2527	0.2604	0.0099	**0.0096**	0.0322	0.0238
		1.023	2.337	0.4041	**0.4040**	0.4046	**0.2543**	0.2571	0.2560	0.2556	0.2621	**0.0156**	0.0160	0.0365	0.0245
	200	2.118	6.996	0.2867	**0.1912**	0.2875	**0.2518**	0.2540	0.2536	0.2536	0.2586	0.0154	**0.0101**	0.0344	0.0237
		1.969	4.997	**0.2896**	0.2903	0.2904	**0.2523**	0.2544	0.2537	0.2537	0.2586	**0.0074**	0.0081	0.0286	0.0236
		1.015	2.863	0.4092	**0.4088**	0.4091	**0.2553**	0.2577	0.2569	0.2567	0.2604	**0.0148**	0.0155	0.0365	0.0244

MLE for fixed effects and random effects variances are consistent and asymptotically normal even if the random effects distribution is misspecified; for instance (Verbeke and Lesaffre, 1997; McCulloch and Neuhaus, 2011). The simulation studies suggest the LMFF performance is robust to the misspecification of the random effects distribution. Formal justification of the robustness for commonly encountered situations needs to be conducted.

In data analysis, we choose the number of eigenfunctions by looking at the cumulative variance explained. As observed in the simulations, large numbers can produce high MSE of estimating β while reducing the fitting errors. As we build a joint model for the predictor and the response, joint model based criteria to choose the numbers may have better performance and are being studied.

We build a point wise asymptotic 95% confidence interval of $\hat{\beta}$ where the standard deviation is obtained via bootstrap. The confidence interval provides a reasonable overall coverage when we simulate data from the fitted model. The coverage proportion varies with regions from 89.9% to 96.8%. Further research will be conducted to achieve more uniform coverage close to the confidence level.

Throughout the analysis, we use an unstructured covariance matrix for the random effects and an independence covariance matrix for the error. In practice, the error can be strongly correlated. It may be more appropriate to fit a model with correlated covariance structures such as compound symmetry or those arising from time series, subject to model identifiability. In standard linear mixed models, model based criteria or graphical plots are used to select the covariance structures for the random effects or for the error; for instance, (Wolfinger, 1993). Similar tools may be derived for the structure selection or model diagnostics.

7. Supplementary Materials

Web Appendices, Tables, and Figures referenced in Sections 2, 3, 4 and 5, and the MATLAB code implementing the new method are available with this paper at the *Biometrics* website on Wiley Online Library.

Acknowledgements

The author would like to thank two referees, the associate editor, and the editor for the speed of review. The author is grateful to their remarks and suggestions that substantially improved the article.

References

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13**, 571–591.

Crainiceanu, C., Reiss, P., Goldsmith, J., Huang, L., Huo, L., Scheipl, F., Swihart, B., Greven, S., Harezlak, J., Kundu, M. G., Zhao, Y., Mclean, M., and Xiao, L. (2013). *refund: Regression with Functional Data*. R package version 0.1-9.

Crainiceanu, C. M., Staicu, A.-M., and Di, C. Z. (2009). Generalized multilevel functional regression. *Journal of American Statistical Association* **104**, 1550–1561.

Crambes, C., Kneip, A., and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Annals of Statistics* **37**, 35–72.

Ferraty, F., Keilegom, I. V., and Vieu, P. (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis* **109**, 10–28.

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2011). Kernel regression with functional response. *Electronic Journal of Statistics* **5**, 159–171.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. New York, NY: Springer.

Garland, T., Jr. (2003). Selection experiments: An under-utilized tool in biomechanics and organismal biology. In *Vertebrate Biomechanics and Evolution*, Bels, V. L., Gasc, J. P., and Casinos, A. (eds), 23–56. Oxford, UK: BIOS Scientific Publishers.

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B. S., and Reich, D. (2011). Penalized functional regression. *Journal of Computational and Graphical Statistics* **20**, 830–851.

Hall, P., Müller, H.-G., and Yao, F. (2008). Modelling sparse generalized longitudinal observations with latent gaussian processes. *Journal of the Royal Statistical Society, Series B* **70**, 703–723.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Liu, C. H. and Rubin, D. B. (1994). The ecme algorithm—A simple extension of em and ecm with faster monotone convergence. *Biometrika* **81**, 633–648.

McCulloch, C. E. and Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science* **26**, 388–402.

Meng, X. L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika* **80**, 267–278.

Morgan, T. J., T. Garland, J., and Carter, P. A. (2003). Ontogenetic trajectories in mice selected for high wheel-running activity. i. mean ontogenetic trajectories. *Evolution* **57**, 646–657.

Peng, J. and Paul, D. (2009). A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics* **18**, 995–1015.

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB (Use R!)*. Springer Series in Statistics.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics, 2nd edition.

Ramsay, J. O., Wickham, H., Graves, S., and Hooker, G. (2013). *fda: Functional Data Analysis*. R package version 2.4.0.

Randolph, T. W., Harezlak, J., and Feng, Z. (2012). Structured penalties for functional linear models—Partially empirical eigenvectors for regression. *Electronic Journal of Statistics* **6**, 323–353.

Reiss, P. T. and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *Journal of American Statistical Association* **102**, 984–996.

Scheipl, F., Staicu, A. M., and Greven, S. (2013). Functional additive mixed models. arXiv:stat/1207.5947v5.

Swallow, J. G., Carter, P. A., and Garland, T., Jr. (1998). Artificial selection for increased wheel-running behavior in house mice. *Behavior Genetics* **28**, 227–237.

Swallow, J. G., Koteja, P., Carter, P. A., and Garland, T., Jr. (1999). Artificial selection for increased wheel-running activity in house mice results in decreased body mass at maturity. *Journal of Experimental Biology* **202**, 2513–2520.

The MathWorks Inc. (2013). *MATLAB*. Natick, Massachusetts, USA.

Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* **23**, 541–556.

Wang, W. (2013). Identifiability of linear mixed effects models. *Electronic Journal of Statistics* **7**, 244–263.

Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in Statistics: Simulation and Computation* **22**, 1079–1106.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics* **33**, 2873–2903.

Received June 2013. *Revised May* 2014. *Accepted May* 2014.