Spline Smoothing For Bivariate Data With Applications To Association Between Hormones

Yuedong Wang *
University of California, Santa Barbara, USA

Wensheng Guo
University of Pennsylvania, Philadelphia, USA

Morton B. Brown
University of Michigan, Ann Arbor, U.S.A.

April 20, 1999

* *Address for correspondence:* Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110. yuedong@pstat.ucsb.edu, phone: (805) 893-4870, Fax: (805) 893-2334.

Abstract

In this paper penalized weighted least-squares is used to jointly estimate nonparametric functions from contemporaneously correlated data. Under conditions generally encountered in practice, it is shown that these joint estimates have smaller posterior variances than those of marginal estimates and are therefore more efficient. We describe three methods: generalized maximum likelihood (GML), generalized cross validation (GCV) and leaving-out-one-pair cross validation (CV) to estimate the smoothing parameters, the weighting parameter and the correlation parameter simultaneously. Based on simulations we conclude that the GML method has smaller mean-square errors for the nonparametric functions and the parameters and needs less computational time than the other methods. Also, it does not overfit data when the sample size is small. Our research is motivated by and is applied to the problem of estimating associations between hormones. We find that the circadian rhythms of the hormones ACTH and cortisol have similar patterns and that cortisol lags ACTH.

Keywords: Circadian rhythms; Cross validation; Generalized cross validation; Generalized maximum likelihood; Nonparametric regression; Penalized weighted least-squares; Seemingly unrelated regression; Self-modeling.

1 Introduction

In many applications two or more dependent variables are observed at several values of the independent variables, such as at multiple time points. The statistical problem is (i) to estimate functions that model their dependences on the independent variables and (ii) to investigate relationships between these functions. This paper was motivated by the problem of possible association between hormone levels that vary over time. Hormone concentrations are assayed in a sequence of blood samples from each of several subjects. The biomedical investigators want to test whether the secretion of one hormone influences the secretion of another and/or whether the concentration level of a hormone provides feedback to another. Even though temporal associations between the two responses do not prove a causal relationship, they provide valuable insights and interesting hypotheses that motivate further studies.

There is a vast literature on parametric regression models with more than one response variable. Zellner (1962) and Gallant (1975) proposed generalized least-squares estimators for a set of seemingly unrelated linear and nonlinear regression functions respectively. Box and Draper (1965) developed a determinant criterion based on a Bayesian argument for multiresponse data. They showed that combining information from several responses provides more precise estimates of the parameters when design matrices are different.

When it is difficult to assume specific forms for the regression functions, it is preferable to make few assumptions about these functions (see, e.g., Wang and Brown (1996)). Nonparametric regression models, especially smoothing splines, provide powerful tools to model these functions. Many authors have considered nonparametric models for multiresponse data. Wegman (1981), Miller and Wegman (1987) and Fessler (1991) proposed algorithms for spline smoothing. Wahba (1992) developed the theory of general smoothing splines using reproducing kernel Hilbert spaces. Soo and Bates (1996) used regression splines. Yee and

Wild (1996) considered additive models for multiresponse data from exponential families. All but Soo and Bates (1996) assumed that the covariance matrix is known, which is usually not true in practice. When the covariance matrix is known, a simple transformation can be used. Therefore the existing methods can be easily modified for calculation. When the covariance matrix is unknown, it has to be estimated from the data and can affect the estimates of the smoothing parameters (Wang 1998).

In this paper the covariance parameters are unknown. Penalized weighted least-squares is used to estimate the nonparametric functions. Contrary to the well-known results for multivariate linear regression models, it is shown that these penalized weighted least-squares estimates are more efficient than function-by-function estimates when design points are the same. The equations for the coefficients in the representations of the solutions are different from those in the independent cases. Therefore, existing algorithms can not be used directly. We propose a method to calculate the estimates and develop the necessary algorithm. We extend the GML and GCV methods to estimate the covariance parameters and the smoothing parameters simultaneously. We also propose a new leaving-out-one-pair CV method and derive a "leaving-out-one-pair" lemma for computations. We compare the GML, GCV and CV methods by simulation.

Kalman filter representations of polynomial smoothing splines were used in Anderson, Jones and Swanson (1990) to model bivariate responses. Their methods were limited to cases where the domains of the functions are continuous intervals. In this paper we consider general spline models in reproducing kernel Hilbert spaces. In doing so, our models can be easily adapted to situations where the domains of the functions are Euclidean spaces and/or observations are linear functionals instead of evaluations. Myers (1991) discussed these cases with examples in earth science. He commented that co-kriging, instead of thin plate spline, was used for multiple responses since "kriging is more easily extended in a natural way". Our proposed methods show that extensions of the spline models are just as easy and natural. Another interesting application of our model is the problem of combining data from multiple sources (Myers 1991, Gao 1994). This problem arises in meteorology where one attribute such as rainfall is measured by several different methods such as rain gauges and radar. Our model can combine these measurements to provide more precise estimates or predictions. For simplicity we use cubic splines in this paper. Our models and methods can be written in terms of general spline models and therefore have many more potential applications.

In the next section, we formulate the model for bivariate responses and propose methods for estimating the nonparametric functions and the parameters in the model. We also show that the joint estimates are more efficient than function-by-function estimates, a property different from that in multivariate linear regression models. In section 3, we conduct simulations to compare the performance of the CV, GCV and GML methods. We fit a SEMOR (self-modeling) model to data consisting of sequences of hormone concentration levels and investigate possible associations between two hormones in section 4. Discussion and conclusions are in section 5.

2 Models and Estimation

2.1 Bivariate Models

For simplicity we only consider bivariate responses. Our methods can be easily extended to multiresponse data. Assume the following model:

$$y_{ki} = f_k(t_{ki}) + \epsilon_{ki}, \qquad k = 1, 2;\ i = 1, \cdots, n_k, \tag{1}$$

where the ith response of the kth variable y_{ki} is generated by the kth function f_k evaluated at the design point t_{ki} plus a random error ϵ_{ki}. Assume $\epsilon_{ki} \overset{iid}{\sim} \mathrm{N}(0, \sigma_k^2)$ for fixed $k = 1, 2$, and $\mathrm{Corr}(\epsilon_{1i}, \epsilon_{2j}) = \rho$ if y_{1i} and y_{2j} are a pair (e.g. observations obtained at the same time) and zero otherwise. We do not assume that the design points t_{1i} and t_{2i} are the same. For a design point where only one response variable is measured, it is conceptually equivalent to a pair with the other response variable being missing at random.

For simplicity of notation, we assume that the domains of both functions are $[0, 1]$ and $f_k \in W_2$, where

$$W_2 = \{f : f, f' \text{ absolutely continuous}, \quad \int_0^1 (f''(t))^2 dt < \infty\}.$$

Our methods can be easily extended to the general smoothing spline models where the two domains are arbitrary (thus could be different) and the observations are linear functionals instead of evaluations (Wahba 1990, Wahba 1992).

2.2 Estimation of the Nonparametric Functions

Denote $\boldsymbol{t}_k = (t_{k1}, \cdots, t_{kn_k})^T$, $\boldsymbol{f}_k = (f_k(t_{k1}), \cdots, f_k(t_{kn_k}))^T$, $\boldsymbol{y}_k = (y_{k1}, \cdots, y_{kn_k})^T$, $\boldsymbol{\epsilon}_k = (\epsilon_{k1}, \cdots, \epsilon_{kn_k})^T$, $\boldsymbol{f} = (\boldsymbol{f}_1^T, \boldsymbol{f}_2^T)^T$, and $\boldsymbol{y} = (\boldsymbol{y}_1^T, \boldsymbol{y}_2^T)^T$, where the superscript ^{T} refers to transpose. Let $\theta = \sigma_1\sigma_2$; $r = \sigma_1/\sigma_2$; J be a $n_1 \times n_2$ matrix with (i, j)th elements equals to 1 if the ith element of $\boldsymbol{y}_1$ and the jth element of $\boldsymbol{y}_2$ is a pair, and zero otherwise. Note that $J = I$, the identity matrix, when all observations come in pairs. Let

$$W^{-1} = \begin{pmatrix} rI_{n_1} & \rho J \\ \rho J^T & I_{n_2}/r \end{pmatrix}, \tag{2}$$

then $\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{f}, \theta W^{-1})$.

The functions f_k are estimated by minimizing the following penalized weighted least-squares

$$\min_{f_1, f_2 \in W_2} \left\{ (\boldsymbol{y} - \boldsymbol{f})^T W (\boldsymbol{y} - \boldsymbol{f}) + \lambda_1 \int_0^1 (f_1''(t))^2 dt + \lambda_2 \int_0^1 (f_2''(t))^2 dt \right\}, \tag{3}$$

where the first part is the weighted least-squares. The remaining parts are penalties for the roughness of the functions. The parameters λ_k control the trade-off between goodness-of-fit and the smoothness of the estimates and are referred to as smoothing parameters.

Let $\phi_1(t) = 1$, $\phi_2(t) = t - 1/2$, $R^1(s, t) = k_2(s)k_2(t) - k_4(s - t)$, where $k_\nu(\cdot) = B_\nu(\cdot)/\nu!$ and $B_\nu(\cdot)$ is the νth Bernoulli polynomial. Let $T_k = \{\phi_\nu(t_{ki})\}_{i=1}^{n_k}{}_{\nu=1}^{2}$; $T = \mathrm{diag}(T_1, T_2)$;

$\Sigma_k = \{R^1(t_{ki}, t_{kj})\}_{i=1}^{n_k}{}_{j=1}^{n_k}$ and $\Sigma = \text{diag}(\Sigma_1, \Sigma_2)$. Similar to Wahba (1990) we can show that for fixed λ_1, λ_2, r and ρ, the solution to (3) is

$$\hat{f}_k(t) = \sum_{\nu=1}^{2} d_{k\nu}\phi_\nu(t) + \sum_{i=1}^{n_k} c_{ki} R^1(t, t_{ki}), \qquad k = 1, 2, \tag{4}$$

where $\boldsymbol{c} = (c_{11}, \cdots, c_{1n_1}, c_{21}, \cdots, c_{2n_2})^T$ and $\boldsymbol{d} = (d_{11}, d_{12}, d_{21}, d_{22})^T$ are solutions to

$$\begin{pmatrix} T^T W T & T^T W \Sigma \\ \Sigma W T & \Sigma W \Sigma + \text{diag}(\lambda_1 \Sigma_1, \lambda_2 \Sigma_2) \end{pmatrix} \begin{pmatrix} \boldsymbol{d} \\ \boldsymbol{c} \end{pmatrix} = \begin{pmatrix} T^T W \boldsymbol{y} \\ \Sigma W \boldsymbol{y} \end{pmatrix}. \tag{5}$$

It is easy to show that $\hat{\boldsymbol{f}} = (\hat{f}_1(t_{11}), \cdots, \hat{f}_1(t_{1n_1}), \hat{f}_2(t_{21}), \cdots, \hat{f}_2(t_{2n_2}))^T = T\boldsymbol{d} + \Sigma\boldsymbol{c}$ is always unique when T is of full column rank, which we assume to be true in this paper. It can be verified that a solution to

$$\begin{cases} (\Sigma + W^{-1}\text{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2}))\boldsymbol{c} + T\boldsymbol{d} &= \boldsymbol{y}, \\ T^T \boldsymbol{c} &= 0, \end{cases} \tag{6}$$

is also a solution to (5). Thus we need to solve (6) for $\boldsymbol{c}$ and $\boldsymbol{d}$. Since observations are correlated, equations (6) are different from (8) in Gu (1989). Therefore existing programs such as RKPACK can not be used to solve (6). In fact, $W^{-1}\text{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2})$ is asymmetric if $\lambda_1 \neq \lambda_2$ and $\rho \neq 0$. To calculate the coefficients $\boldsymbol{c}$ and $\boldsymbol{d}$, we use the following transformations: $\tilde{\Sigma} = \Sigma\text{diag}(I_{n_1}/\lambda_1, I_{n_2}/\lambda_2)$ and $\tilde{\boldsymbol{c}} = \text{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2})\boldsymbol{c}$. Then (6) is equivalent to

$$\begin{cases} (\tilde{\Sigma} + W^{-1})\tilde{\boldsymbol{c}} + T\boldsymbol{d} &= \boldsymbol{y}, \\ T^T \tilde{\boldsymbol{c}} &= 0. \end{cases} \tag{7}$$

Let

$$T_k = (Q_{k1}, Q_{k2}) \begin{pmatrix} R_k \\ 0 \end{pmatrix}, \qquad k = 1, 2,$$

be the QR decompositions. Let $Q_1 = \text{diag}(Q_{11}, Q_{21})$; $Q_2 = \text{diag}(Q_{12}, Q_{22})$; $R = \text{diag}(R_1, R_2)$ and $B = \tilde{\Sigma} + W^{-1}$. Similar to Wahba (1990), it can be shown that the solutions to (7) are

$$\begin{aligned} \tilde{\boldsymbol{c}} &= Q_2(Q_2^T B Q_2)^{-1} Q_2^T \boldsymbol{y}, \\ R\boldsymbol{d} &= Q_1^T(\boldsymbol{y} - B\tilde{\boldsymbol{c}}). \end{aligned} \tag{8}$$

It is easy to check that $\hat{\boldsymbol{f}} = A\boldsymbol{y}$, where

$$A = I - W^{-1} Q_2 (Q_2^T B Q_2)^{-1} Q_2^T \tag{9}$$

is the "hat" matrix. Note that A is not symmetric, which is different from the usual independent case.

2.3 Estimations of Parameters

So far we have assumed that the parameters λ_1, λ_2, r and ρ are fixed. In practice it is very important to estimate these parameters from the data. Since observations are correlated, popular methods such as the usual GML method and the GCV method may underestimate the smoothing parameters λ_1 and λ_2 (Wang 1998). In this section we propose the following three methods to estimate the smoothing parameters λ_1 and λ_2, the weighting parameter r and the correlation parameter ρ simultaneously: (1) an extension of the GML method based on a Bayesian model; (2) an extension of the GCV method; and (3) leaving-out-one-pair cross validation. We compare these three methods through simulation in section 3.

Wang (1998) proposed the GML and GCV methods for correlated observations with one smoothing parameter. In a bivariate model, there are two smoothing parameters which need to be estimated simultaneously together with the covariance parameters. Following a similar derivation, we extend the GML and GCV in Wang (1998) as follows.

The GML estimates of λ_1, λ_2, r and ρ are minimizers of the following GML function:

$$M(\lambda_1, \lambda_2, r, \rho) = \frac{\boldsymbol{y}^T W(I-A)\boldsymbol{y}}{[\det^+(W(I-A))]^{\frac{1}{n-4}}} = \frac{\boldsymbol{z}^T (Q_2^T B Q_2)^{-1}\boldsymbol{z}}{[\det(Q_2^T B Q_2)^{-1}]^{\frac{1}{n-4}}}\ , \tag{10}$$

where $n = n_1 + n_2$; $\det^+$ is the product of the nonzero eigenvalues and $\boldsymbol{z} = Q_2^T \boldsymbol{y}$. The minimizers of $M(\lambda_1, \lambda_2, r, \rho)$ are called GML estimates.

The GCV estimates of λ_1, λ_2, r and ρ are minimizers of the following GCV function:

$$V(\lambda_1, \lambda_2, r, \rho) = \frac{||W(I-A)\boldsymbol{y}||^2}{[\mathrm{Tr}(W(I-A))]^2} = \frac{\boldsymbol{z}^T (Q_2^T B Q_2)^{-2}\boldsymbol{z}}{[\mathrm{Tr}(Q_2^T B Q_2)^{-1}]^2}\ . \tag{11}$$

In the following we propose a cross validation method based on a leaving-out-one-pair procedure. Suppose there are a total of N ($N \geq \max\{n_1, n_2\}$) distinct time points and thus N pairs of observations. Any one observation in a pair may be missing. These pairs are numbered from 1 to N. We use the following notation: superscripts (i) to denote the collection of elements corresponding to the ith pair; superscripts $[i]$ to denote the collection of elements after deleting the ith pair; superscripts $\{i\}$ to denote solution of f_k without the ith pair. When one observation in a pair is missing, superscripts indicate a single observation instead of a pair.

$\hat{f}_1^{\{i\}}$ and $\hat{f}_2^{\{i\}}$ are solutions to

$$\min_{f_1, f_2 \in W_2} \left\{ (\boldsymbol{y}^{[i]} - \boldsymbol{f}^{[i]})^T W^{[i]} (\boldsymbol{y}^{[i]} - \boldsymbol{f}^{[i]}) + \lambda_1 \int_0^1 (f_1''(t))^2 dt + \lambda_2 \int_0^1 (f_2''(t))^2 dt \right\}. \tag{12}$$

Assume that there are two elements in the ith pair (it is simple if there is only one). Denote i_1 and i_2 as the row numbers of this pair in $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ respectively. Define

$$y_{kj}^* = \begin{cases} y_{kj}, & j \neq i_k \\ \hat{f}_k^{\{i\}}(t_{ki_k}), & j = i_k, \end{cases} \qquad k = 1, 2.$$

Denote $\boldsymbol{y}_k^* = (y_{k1}^*, \cdots, y_{kn_k}^*)^T$, $\boldsymbol{y}^* = (\boldsymbol{y}_1^{*T}, \boldsymbol{y}_2^{*T})^T$ and $\hat{f}^{\{i\}}(\boldsymbol{t}) = (\hat{f}_1^{\{i\}}(t_{11}), \cdots, \hat{f}_1^{\{i\}}(t_{1n_1}),$ $\hat{f}_2^{\{i\}}(t_{21}), \cdots, \hat{f}_2^{\{i\}}(t_{2n_2}))$. Then we have the following "leaving-out-one-pair" lemma:

Lemma 1 *For fixed λ_1, λ_2, r, ρ and i, $\hat{f}^{\{i\}}(\boldsymbol{t}) = A\boldsymbol{y}^*$.*

As a consequence of this lemma, we do not need to solve separate minimization problems (12) for each deleting-one-pair set. All we need to do is to solve the following equations

$$\begin{pmatrix} 1-a(i_1,i_1) & -a(i_1,n_1+i_2) \\ -a(n_1+i_2,i_1) & 1-a(n_1+i_2,n_1+i_2) \end{pmatrix} \begin{pmatrix} \hat{f}_1^{\{i\}}(t_{1i_1}) - y_{1i_1} \\ \hat{f}_2^{\{i\}}(t_{2i_2}) - y_{2i_2} \end{pmatrix} = \begin{pmatrix} \hat{f}_1(t_{1i_1}) - y_{1i_1} \\ \hat{f}_2(t_{2i_2}) - y_{2i_2} \end{pmatrix} \quad (13)$$

for $\hat{f}_k^{\{i\}}(t_{ki_k}) - y_{ki_k}$, where $a(i,j)$ are elements of the matrix A. If there is only one observation in the ith pair, for example y_{1i_1}, we then have the following equation

$$(1-a(i_1,i_1))(\hat{f}_1^{\{i\}}(t_{1i_1}) - y_{1i_1}) = \hat{f}_1(t_{1i_1}) - y_{1i_1}. \quad (14)$$

Note that (14) is exactly the same as the "leaving-out-one" lemma in the independent case. The proofs of Lemma 1 and (13) are given in the Appendix.

Let $\hat{\boldsymbol{f}}_k^{\{-\}} = (\hat{f}_k^{\{i_{k1}\}}(t_{k1}), \cdots, \hat{f}_k^{\{i_{kn_k}\}}(t_{kn_k}))^T$ and $\hat{\boldsymbol{f}}^{\{-\}} = (\hat{\boldsymbol{f}}_1^{\{-\}^T}, \hat{\boldsymbol{f}}_2^{\{-\}^T})^T$, where i_{kj} denotes the index of the pair for observation y_{kj}. Define the cross validation score as

$$C(\lambda_1, \lambda_2, r, \rho) = \frac{1}{n}||W(\boldsymbol{y} - \hat{\boldsymbol{f}}^{\{-\}})||^2. \quad (15)$$

C estimates the weighted mean-square errors (WMSE) (Wang 1998). The minimizers of $C(\lambda_1, \lambda_2, r, \rho)$ are called cross validation estimates of the parameters.

It is obvious from the proofs in the Appendix that the leaving-out-one-pair lemma is true for leaving-out-one-cluster. So this cross validation method can be easily modified for longitudinal data and curved data, where a cluster is a subject or a curve. This lemma is more general than that of Rice and Silverman (1991) since weighted least-squares, instead of least-squares is used and we do not assume that the design points are the same for all curves.

We developed a new package BIVSPLINE using Fortran 77 and some subroutines from LINPACK (Dongarra, Moler, Bunch and Stewart 1979). This package solves equation (7) and estimates smoothing parameters and covariance parameters using GML, GCV and leaving-out-one-pair CV. This package is available from the authors.

2.4 Properties of the joint estimates

In section 2.2 we proposed to estimate functions f_1 and f_2 jointly by minimizing the joint weighted least-squares. One may also estimate these functions separately by minimizing the marginal least-squares. For multivariate linear regression models it is well-known that the joint estimates equal function-by-function estimates when the sampling points for the two functions are the same ($\boldsymbol{t}_1 = \boldsymbol{t}_2$). When sampling points are different, the joint estimates are more efficient (Zellner 1962, Gallant 1975). In the smoothing spline context the posterior covariances are often used to construct Bayesian confidence intervals (Wahba 1990). In this section we show that the covariance matrix of the separately estimated functions minus the covariance matrix of the jointly estimated functions is positive semi-definite when $\boldsymbol{t}_1 = \boldsymbol{t}_2$. Therefore the joint estimates are more efficient even when design points are the same.

Consider the following prior for f_k:

$$F_k(t) = \sum_{\nu=1}^{2} \delta_{k\nu}\phi_\nu(t) + b_k^{1/2} X_k(t), \qquad k = 1, 2, \tag{16}$$

where $\delta_{k\nu} \overset{iid}{\sim} \mathrm{N}(0, a)$; a and b_k are positive constants; $X_k(t)$ is a zero mean Gaussian stochastic process independent of $\delta_{k\nu}$ with covariance function $\mathrm{E}(X_k(s)X_k(t)) = R^1(s,t)$. Assume F_1 and F_2 are independent and observations are generated by

$$\boldsymbol{y}_k = \boldsymbol{F}_k + \boldsymbol{\epsilon}_k, \qquad k = 1, 2, \tag{17}$$

where $\boldsymbol{F}_k = (F_k(t_{k1}), \cdots, F_k(t_{kn_k}))^T$, $\boldsymbol{\epsilon}_k$ is as defined in (1). Let $\boldsymbol{F} = (\boldsymbol{F}_1^T, \boldsymbol{F}_2^T)^T$. Following the same arguments as in Wahba (1990) and Wang (1998), with $\lambda_k = \theta/b_k$, we have

$$\begin{aligned} &\lim_{a\to\infty} \mathrm{E}(\boldsymbol{F}|\boldsymbol{y}) = \hat{\boldsymbol{f}}, \\ &\lim_{a\to\infty} \mathrm{Cov}(\boldsymbol{F}|\boldsymbol{y}) = \theta A W^{-1}. \end{aligned} \tag{18}$$

For simplicity we assume that there are no missing data and $\boldsymbol{t}_1 = \boldsymbol{t}_2$. Let $\tilde{f}_k$ be the function-by-function estimates. That is, they are minimizers of the following separate marginal penalized least-squares

$$\min_{f_k \in W_2} \left\{ (\boldsymbol{y}_k - \boldsymbol{f}_k)^T(\boldsymbol{y}_k - \boldsymbol{f}_k) + \tilde{\lambda}_k \int_0^1 (f_k''(t))^2 dt \right\}, \tag{19}$$

where $\tilde{\lambda}_k$ are smoothing parameters. Let $\tilde{\boldsymbol{f}}_k = (\tilde{f}_k(t_{k1}), \cdots, \tilde{f}_k(t_{kn_k}))$ and $\tilde{\boldsymbol{f}} = (\tilde{\boldsymbol{f}}_1^T, \tilde{\boldsymbol{f}}_2^T)^T$. From Wahba (1990), $\tilde{\boldsymbol{f}}_k = A_k \boldsymbol{y}_k$, where $A_k = I - Q_{k2}(Q_{k2}^T(I + \Sigma_k/\tilde{\lambda}_k)Q_{k2})^{-1}Q_{k2}^T$. With the same Bayesian model defined in (16) and (17), it is shown in the Appendix that with $\tilde{\lambda}_k = \sigma_k^2/b_k$,

$$\begin{aligned} \lim_{a\to\infty} \mathrm{E}(\boldsymbol{F}_k|\boldsymbol{y}_k) &= \tilde{\boldsymbol{f}}_k, \\ \lim_{a\to\infty} \mathrm{Cov}(\tilde{\boldsymbol{f}} - \boldsymbol{F}) &= \theta G, \end{aligned} \tag{20}$$

where

$$G = \begin{pmatrix} rA_1 & \rho A_1 A_2 \\ \rho A_2 A_1 & A_2/r \end{pmatrix}.$$

Theorem 1 *$G - AW^{-1}$ is positive semi-definite. Furthermore, if $\rho \neq 0$, $\lambda_1 < \infty$, $\lambda_2 < \infty$ and Σ_k are full rank, then $\boldsymbol{x}^T(G - AW^{-1})\boldsymbol{x} > 0$ for any $\boldsymbol{x} \bar{\in} span(T)$.*

The proof is given in the Appendix. The diagonal elements of these posterior covariances are often used to construct Bayesian confidence intervals. Theorem 1 indicates that the Bayesian confidence intervals of the joint estimates are narrower than those of the separate estimates. Since vectors such as all elements except one equal to zero are usually not in the space spanned by T, the confidence intervals of the joint estimates are usually strictly narrower. Note that when $\lambda_1 = \infty$ and $\lambda_2 = \infty$, the spline models are reduced to multivariate linear regression models. As is well-known, the joint estimates and the function-by-function estimates are the same.

3 Simulation

In this section we conduct simulations to compare three methods: CV, GCV and GML to estimate the parameters and to evaluate the estimates of functions. The underlying model is

$$
\begin{aligned}
y_{1i} &= \sin 2\pi i/N + \epsilon_{1i}, \\
y_{2i} &= \sin 2\pi (i/N)^2 + \epsilon_{2i}, \qquad i = 1, \cdots, N.
\end{aligned}
$$

Four different sample sizes $N = 50, 100, 200, 400$, all four combinations of standard deviations $\sigma_1, \sigma_2 = 0.1, 0.2$, and seven different correlations $\rho = -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75$ are used. Other functions and combinations of parameters are also used. Since all the results are similar, we only report results for $N = 100$, $\sigma_1 = \sigma_2 = 0.1$ and $\rho = 0.5$. Responses are generated for 100 replications.

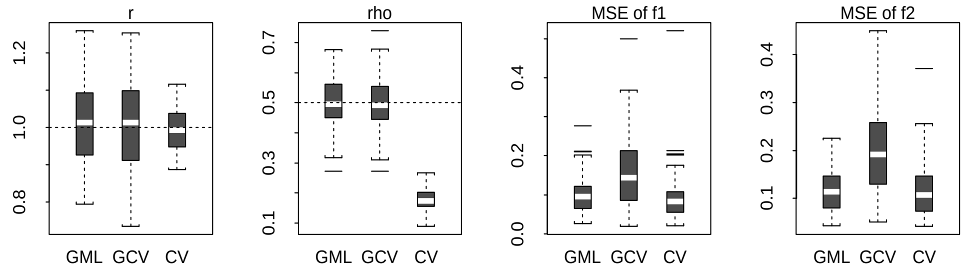

Figure 1: Plots from left to right are: estimates of r, estimates of ρ, mean-square errors of $\hat{f}_1$ and mean-square errors of $\hat{f}_2$. The dotted lines in the left two plots indicate the true values.

Figure 1 presents the estimates of r and ρ, and mean-square errors of $\hat{f}_1$ and $\hat{f}_2$ of the 100 replications. From these plots we can see that the GML and GCV estimates of r and ρ have small biases and variances. The CV estimates of r have smaller variances, but the CV estimates of ρ have larger biases. From the mean-square errors and plots of the estimated functions (not shown here) we conclude that all three methods estimate the smoothing parameters and the functions well. The GML and CV provide better estimates than GCV in terms of mean-square errors. The GML method is more stable when the sample size is small such as $N = 50$. In this case there were several replications where GCV and CV provided very small estimates of smoothing parameters which lead to over-fitting the data. This behavior of GCV method was investigated in Wahba and Wang (1993) and Wang (1998). The GCV method performs as well as the GML method for moderate sample sizes ($N = 200,\ 400$) and better than the GML method for very large sample sizes. The CV method is computationally more intensive than the GML and GCV methods. Overall the GML method works well and is recommended.

To examine the shape of the functions of M, V and C, we show "typical" (i.e., the first replicate of the simulation) contour plots of these three functions in Figure 2. We can see from these plots that each function has an unique minimum inside the search region which is near the true value of r. The GML and GCV functions have minima near the true value of ρ, while the minimum of the CV function is far away. All functions have nice bowl shapes near their local minima.

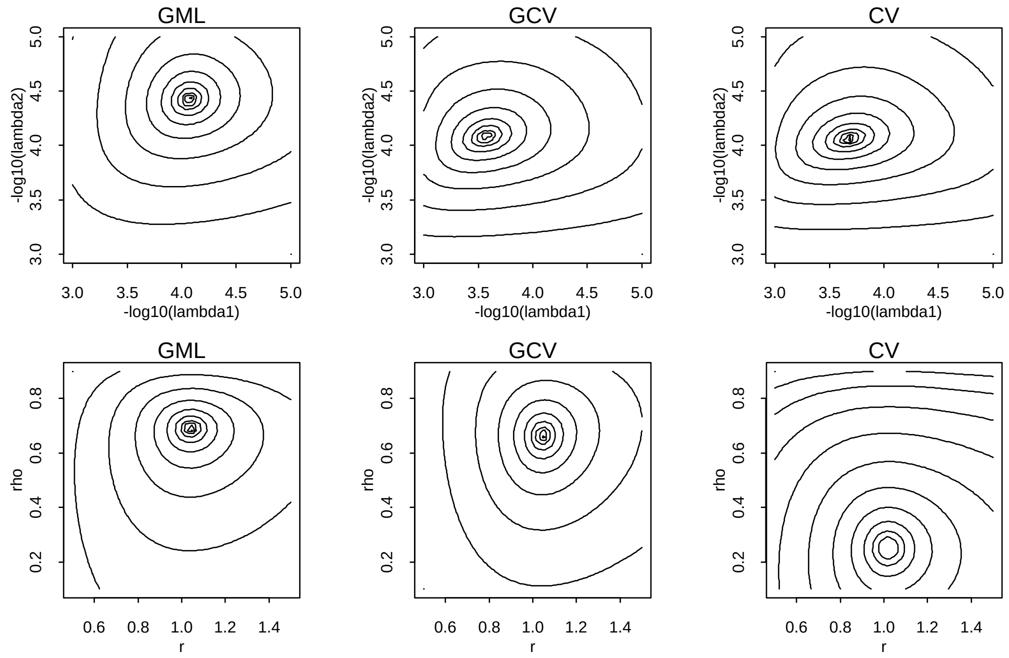

Figure 2: Contour plots of the three functions of the first replicate in the simulation with $N = 100$, $\sigma_1 = \sigma_2 = 0.1$ and $\rho = 0.5$. In the first row criteria scores are plotted as functions of two smoothing parameters with $r = 1$ and $\rho = 0.5$. In the second row criteria scores are plotted as functions of r and ρ with two smoothing parameters fixed at $-\log_{10}\lambda_1 = -\log_{10}\lambda_2 = 4$.

To compare empirically the marginal and the joint methods, we present ratios between mean-square errors of the function-by-function estimators and mean-square errors of the penalized weighted least-squares estimators in Figure 3. Based on this and other simulations not shown here, we find the ratio increases with the increase of correlation, complexity of the function and variances. The improvement in this particular case ($N = 100$, $\sigma_1 = \sigma_2 = 0.1$ and $\rho = 0.5$) is small. For some simulations the improvement is more than 30%.

For marginal estimators, the diagonal elements in (20) are often used to construct Bayesian confidence intervals (Wahba 1983). Similar Bayesian confidence intervals can be constructed for the joint estimators using the diagonal elements in (18). Figure 4 show that the Bayesian confidence intervals for the joint estimators performed well in our simulations.

4 Associations Between Hormones

To illustrate our methods, we use data from a stress study conducted in the medical center of the University of Michigan. Blood samples were collected every 10 minutes for 24 hours from 8 healthy normal female volunteers. These blood samples were assayed for concentrations of

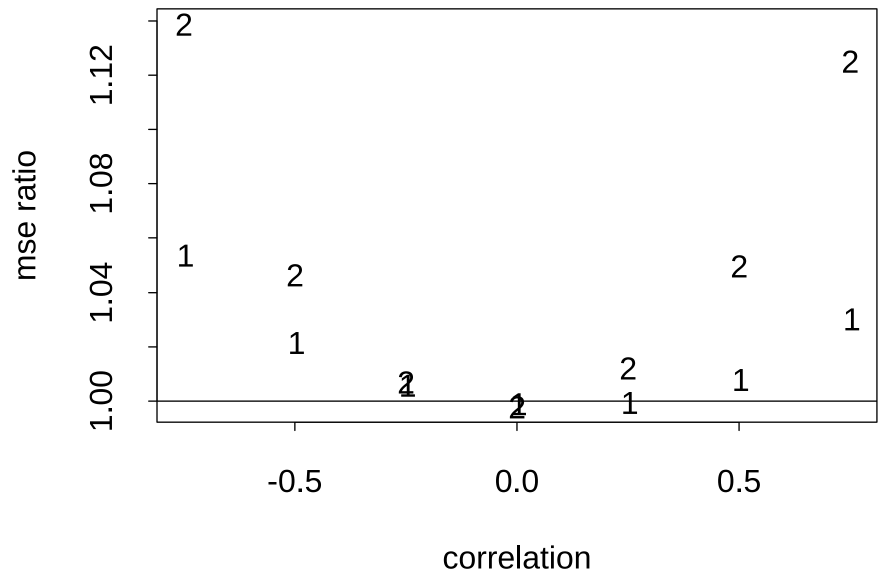

Figure 3: Ratios between mean-square errors of the marginal estimators and mean-square errors of the joint estimators. Points are marked as 1 for the first function and 2 for the second function.

both adrenocortocotropic (ACTH) and cortisol. This experiment was intended to investigate the pulsatile behavior of these hormones. Approval for this experiment was obtained from the institutional human subjects review committee and informed consent was obtained from all participants. In this paper we are primarily interested in modeling circadian rhythms. For each subject we select a subset of the data that contains the assay values at two hours intervals starting at 9:00 am. These subseries mimic experiments usually conducted for investigating circadian rhythms where the sampling rate is low since the long term trend is of interest. We also fitted other subseries and the conclusions remain the same.

Figure 5 presents the observed concentrations of ACTH and cortisol on a log scale. It is obvious from these plots that there is a large variation between subjects. The investigators believe that there is a common function for each hormone that generates the responses. However, the time axis may be shifted and the magnitude of the responses may differ between subjects. Therefore we assume the following SEMOR models for ACTH and cortisol:

$$y_{ijk} = \mu_{ik} + \alpha_{ik} f_k(t_j - \tau_{ik}) + \epsilon_{ijk}, \qquad i = 1, \cdots, 8; \quad j = 1, \cdots, 12; \quad k = 1, 2, \tag{21}$$

where $k = 1$ and $k = 2$ correspond to ACTH and cortisol respectively. y_{ijk} is the hormone k concentration on a log scale of the ith individual at the jth time point t_j, μ_{ik} is the hormone k 24-h mean of the ith individual, α_{ik} is the hormone k amplitude of the ith individual, τ_{ik} is the hormone k phase of the ith individual, f_k is the common function of hormone k,

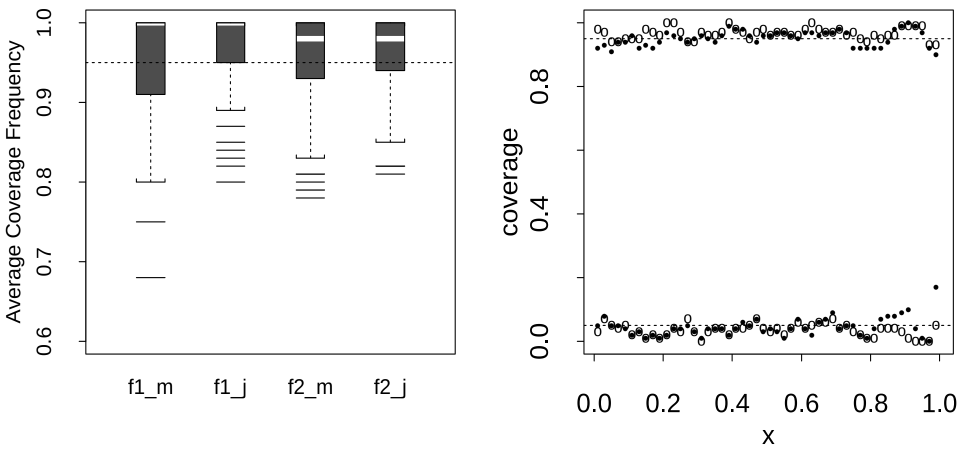

Figure 4: Left: Box plots of the across the function coverage of 95% Bayesian confidence intervals. Labels "f1" and "f2" represent the first and the second functions. "m" and "j" represent marginal and joint estimators. Right: above are the pointwise percentages that the 95% Bayesian confidence intervals cover the true first function; below are the pointwise percentages that the 95% Bayesian confidence intervals miss the true second function. Stars represent the marginal estimators and circles represent the joint estimators. Dotted lines: nominal values.

and ϵ_{ijk}'s are random errors. Since measurements of these two hormones were taken from the same blood samples and pulses are ignored, ϵ_{ijk} may be correlated for the same i and j. Therefore it is better to estimate the functions f_k jointly. We assume that $\epsilon_{ijk} \overset{iid}{\sim} \mathrm{N}(0, \sigma_k^2)$ for fixed $k = 1, 2$, and $\mathrm{Corr}(\epsilon_{ij1}, \epsilon_{ij2}) = \rho$. The common functions f_k are modeled by periodic splines with the constraints:

$$\int_0^1 f_k(t)dt = 0, \qquad \sup_{t\in[0,1]} |f_k(t)| = 1. \tag{22}$$

See Wang and Brown (1996) for details.

We follow the same procedure as that in Wang and Brown (1996) to estimate the parameters μ_{ik}, α_{ik}, τ_{ik} and the common functions f_k, where the common functions are now estimated by the methods proposed in previous sections. The GML method is used to select the smoothing parameters, the weighting parameter and the correlation parameter. Estimates of the parameters σ_1^2, σ_2^2 and ρ are 0.2000 (0.1477,0.2672), 0.1832 (0.1367,0.2454) and 0.3864 (0.1905,0.5745) respectively, where the numbers in the parentheses are their 95% bootstrap confidence intervals. See Wang and Brown (1996) for details about the bootstrap procedure. Since the confidence interval for ρ does not include zero, we conclude that there is a positive correlation in the residuals. Fitted lines are plotted in Figure 5. Define acrophase h as the location of the maximum response. Estimates and bootstrap standard deviations of the 24-h mean μ, amplitude α, and acrophase h for ACTH and cortisol of all subjects and their means are listed in Table 1. On average, ACTH and cortisol have similar amplitudes and peak points.

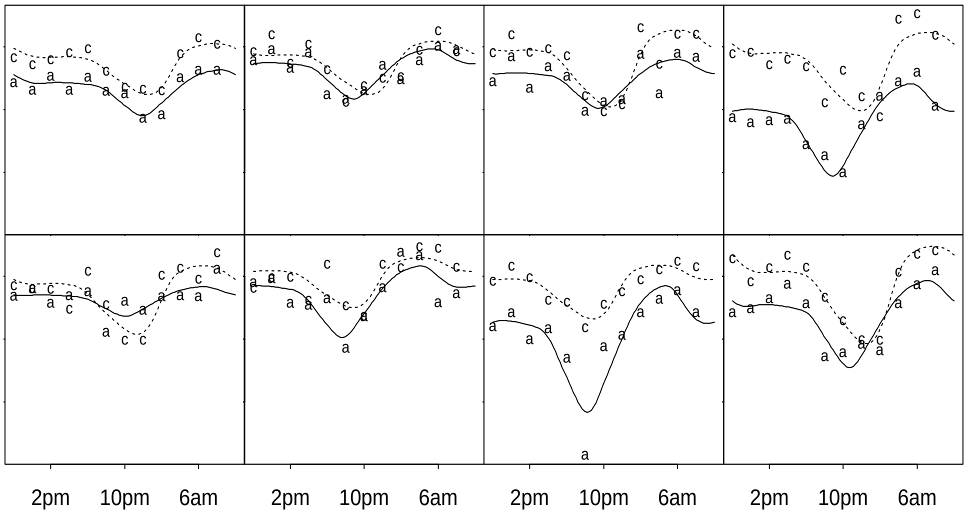

Figure 5: Plots of ACTH and cortisol concentrations on a log scale for eight subjects. Letters "a" represent observations of ACTH and letters "c" represent observations of cortisol. The solid lines are estimates of ACTH and the dotted lines are estimates of cortisol.

subject	1	2	3	4	5	6	7	8	mean
	ACTH								
$\hat{\mu}$	0.69	1.31	0.99	-0.33	1.29	1.43	0.12	0.78	0.71
$\hat{s}(\hat{\mu})$	0.15	0.14	0.13	0.13	0.13	0.13	0.12	0.13	0.13
$\hat{\alpha}$	0.89	0.97	0.95	1.79	0.57	1.38	2.46	1.68	1.33
$\hat{s}(\hat{\alpha})$	0.28	0.25	0.26	0.27	0.23	0.28	0.30	0.29	0.27
$\hat{h}$	7:18	4:23	4:52	4:23	5:36	3:10	3:40	6:05	4:56
$\hat{s}(\hat{h})$	69	66	70	53	116	61	47	46	66
	Cortisol								
$\hat{\mu}$	1.47	1.52	1.58	1.49	1.49	1.97	1.69	1.75	1.61
$\hat{s}(\hat{\mu})$	0.13	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
$\hat{\alpha}$	0.99	1.04	1.50	1.52	1.33	0.97	1.05	1.89	1.28
$\hat{s}(\hat{\alpha})$	0.25	0.23	0.27	0.25	0.24	0.26	0.27	0.26	0.26
$\hat{h}$	6:34	4:52	4:52	6:05	5:21	2:56	2:56	6:34	5:01
$\hat{s}(\hat{h})$	64	68	56	54	56	66	64	54	60

Table 1: Estimates of μ, α, h and their bootstrap standard deviations for log(ACTH) and log(cortisol). $\hat{s}(\hat{h})$ are in minutes.

The investigators are interested in the possible association between ACTH and cortisol. There are many potential associations. One interesting hypothesis is that the two common functions f_1 and f_2 are linearly related: $f_1(t) = a+bf_2(t+c)$. It is equivalent to $f_1(t) = \pm f_2(t)$ because of the constraints (22) and the phases. For ACTH and cortisol, we are interested in positive associations. Therefore we want to test

$$\mathrm{H}_0: \ f_1 = f_2, \qquad \mathrm{H}_1: f_1 \neq f_2. \tag{23}$$

Under the null hypothesis, hormones ACTH and cortisol are generated by the same common function. Thus we can think of the whole data set as one hormone from 16 subjects. The procedure developed in Wang and Brown (1996) can then be used to fit data under the null hypothesis. The estimated common function together with 95% Bayesian confidence intervals are plotted in Figure 6. We superimpose the shifted estimates of the common functions under H_1; that is, they are shifts of the estimates of f_1 and f_2 in model (21). We can see that a large portion of these two shifted functions are inside the Bayesian confidence intervals, which suggests that H_0 is true.

To formally test the null hypothesis, we calculate a F type statistic

$$F = \frac{(\mathrm{RSS}_{\mathrm{H}_0} - \mathrm{RSS}_{\mathrm{H}_1})/(\mathrm{d.f.}_{\mathrm{H}_0} - \mathrm{d.f.}_{\mathrm{H}_1})}{\mathrm{RSS}_{\mathrm{H}_1}/\mathrm{d.f.}_{\mathrm{H}_1}},$$

and a likelihood ratio statistic

$$D = 2 \times (l_{\mathrm{H}_1} - l_{\mathrm{H}_0}),$$

where $\mathrm{RSS}_{\mathrm{H}_0}$ and $\mathrm{RSS}_{\mathrm{H}_1}$ are residual sum of squares under H_0 and H_1 respectively, $\mathrm{d.f.}_{\mathrm{H}_0}$ and $\mathrm{d.f.}_{\mathrm{H}_1}$ are degrees of freedom for residuals under H_0 and H_1 respectively, and l_{H_0} and l_{H_1} are the log likelihoods under H_0 and H_1 respectively. We use a bootstrap procedure to calculate the null distributions of F and D. That is, we fit the data under H_0. We then generate bootstrap samples from this fit. These bootstrap samples are fitted under both H_0 and H_1 to get bootstrap estimates F^* of F and D^* of D under H_0. The empirical distributions of these F^*'s and D^*'s are used as the null distributions. Histograms of the bootstrap statistics F^* and D^* are plotted in Figure 7. The p-values based on F^* and D^* are 0.7847 and 0.8033 respectively. Therefore, we can not reject H_0. Under H_0, that is using the same common function for both ACTH and cortisol, we can calculate the lags between ACTH and cortisol for each subject based on the estimates of the phases. The average lag is 42 minutes, which means that cortisol lags ACTH by 42 minutes on average.

5 Discussion

General smoothing spline models provide flexibility for estimating nonparametric functions and are widely used in many areas. With multiple correlated responses it is better to estimate these functions jointly using the penalized weighted least-squares. Data driven methods for selecting the smoothing parameters are very important for smoothing spline estimates and are quite difficult for correlated observations. We propose to estimate the smoothing parameters and the covariance parameters simultaneously. Our leaving-out-one-pair cross

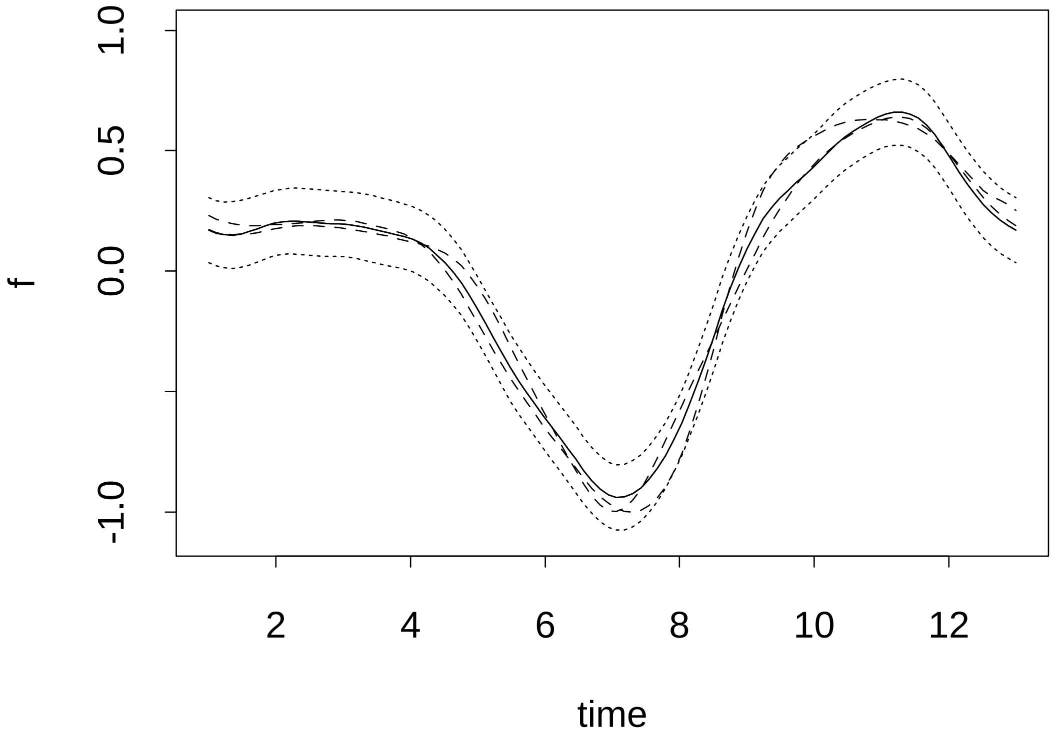

Figure 6: The wide solid line is the estimate of the common function under H_0. The two dotted lines are 95% Bayesian confidence intervals. The two narrow solid lines are the shifted estimates of the common functions under H_1.

validation procedure should be viewed broadly as leaving-out-one-cluster procedure. Therefore it can be used to select the smoothing parameters for longitudinal data and curved data. Missing data often occur in practice and our model deals with this situation naturally.

Biomedical scientists are often interested in two different kinds of associations between hormones: the associations between the baselines such as circadian rhythms and the associations between pulses. We have ignored the pulses since we are interested in the association between circadian rhythms. We will investigate the associations of the baselines and the pulses simultaneously in our future research. Our first step is to develop a method to estimate the baseline and pulses jointly for a single hormone series (Guo, Wang and Brown 1999).

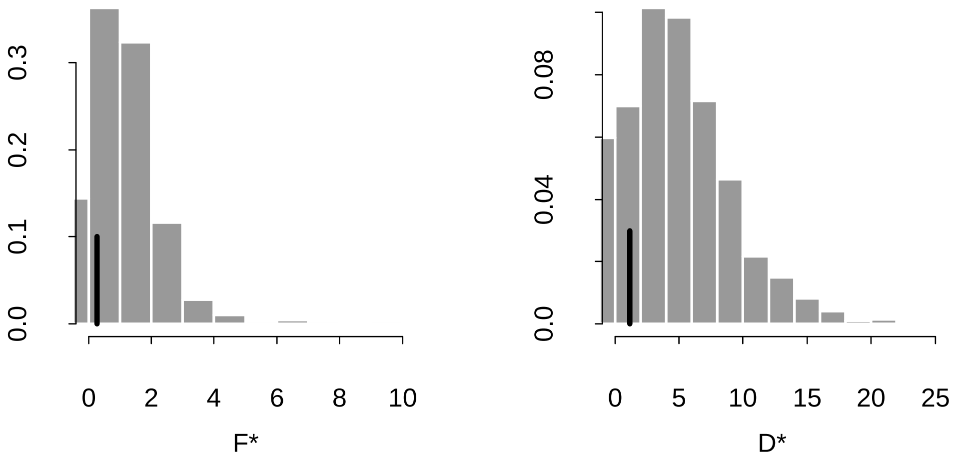

Figure 7: Histograms of F^* and D^*. Vertical bars are test statistics F and D.

ACKNOWLEDGMENTS

This research was supported by NIH Grants R01 GM58533, P60 DK20572, P30 HD18258, M01 RR00042 and R01 EY09946. We would like to thank Dr. Leslie Crofford for allowing us to use the hormone data. We would also like to thank the associate editor for thoughful suggestions that improved the article.

6 Appendix

Proof of Lemma 1

Let $f(\boldsymbol{t}) = (f_1(t_{11}), \cdots, f_1(t_{1n_1}), f_2(t_{21}), \cdots, f_2(t_{2n_2}))$ and $\hat{f}^{\{i\}}(\boldsymbol{t}) = (\hat{f}_1^{\{i\}}(t_{11}), \cdots, \hat{f}_1^{\{i\}}(t_{1n_1}), \hat{f}_2^{\{i\}}(t_{21}), \cdots, \hat{f}_2^{\{i\}}(t_{2n_2}))$. Similarly define $f(\boldsymbol{t}^{[i]})$ and $\hat{f}^{\{i\}}(\boldsymbol{t}^{[i]})$ as $f(\boldsymbol{t})$ and $\hat{f}^{\{i\}}(\boldsymbol{t})$ respectively without the elements corresponding to the ith pair. For any functions f_1 and f_2 in W_2,

$$\begin{aligned} & (\boldsymbol{y}^* - f(\boldsymbol{t}))^T W (\boldsymbol{y}^* - f(\boldsymbol{t})) + \lambda_1 \int_0^1 (f_1''(t))^2 dt + \lambda_2 \int_0^1 (f_2''(t))^2 dt \\ \geq \;& (\boldsymbol{y}^{[i]} - f(\boldsymbol{t}^{[i]}))^T W^{[i]} (\boldsymbol{y}^{[i]} - f(\boldsymbol{t}^{[i]})) + \lambda_1 \int_0^1 (f_1''(t))^2 dt + \lambda_2 \int_0^1 (f_2''(t))^2 dt \\ \geq \;& (\boldsymbol{y}^{[i]} - \hat{f}^{\{i\}}(\boldsymbol{t}^{[i]}))^T W^{[i]} (\boldsymbol{y}^{[i]} - \hat{f}^{\{i\}}(\boldsymbol{t}^{[i]})) + \lambda_1 \int_0^1 [(\hat{f}_1^{\{i\}}(t))'']^2 dt + \lambda_2 \int_0^1 [(\hat{f}_2^{\{i\}}(t))'']^2 dt \\ = \;& (\boldsymbol{y}^* - \hat{f}^{\{i\}}(\boldsymbol{t}))^T W (\boldsymbol{y}^* - \hat{f}^{\{i\}}(\boldsymbol{t})) + \lambda_1 \int_0^1 [(\hat{f}_1^{\{i\}}(t))'']^2 dt + \lambda_2 \int_0^1 [(\hat{f}_2^{\{i\}}(t))'']^2 dt, \end{aligned} \tag{24}$$

where the first inequality holds because after switching rows and columns, we have

$$\begin{aligned} & (\boldsymbol{y}^* - f(\boldsymbol{t}))^T W (\boldsymbol{y}^* - f(\boldsymbol{t})) \\ = \;& \begin{pmatrix} \boldsymbol{y}^{*[i]} - f(\boldsymbol{t}^{[i]}) \\ \boldsymbol{y}^{*(i)} - f(\boldsymbol{t}^{(i)}) \end{pmatrix}^T \begin{pmatrix} W^{[i]} & 0 \\ 0 & W^{(i)} \end{pmatrix} \begin{pmatrix} \boldsymbol{y}^{*[i]} - f(\boldsymbol{t}^{[i]}) \\ \boldsymbol{y}^{*(i)} - f(\boldsymbol{t}^{(i)}) \end{pmatrix} \end{aligned}$$

$$\geq \ (\boldsymbol{y}^{[i]} - f(\boldsymbol{t}^{[i]}))^T W^{[i]} (\boldsymbol{y}^{[i]} - f(\boldsymbol{t}^{[i]})).$$

The second inequality holds because $\hat{f}_1^{\{i\}}$ and $\hat{f}_2^{\{i\}}$ are solution to (12). The last equality holds because of the definition of $\boldsymbol{y}^*$.

(24) indicates that $\hat{f}_1^{\{i\}}$ and $\hat{f}_2^{\{i\}}$ are solutions to (3) with $\boldsymbol{y}$ replaced by $\boldsymbol{y}^*$. Therefore $\hat{f}^{\{i\}}(\boldsymbol{t}) = A\boldsymbol{y}^*$.

Proof of equation (13)

From Lemma 1, we have

$$\begin{aligned}
& \hat{f}_1^{\{i\}}(t_{1i_1}) - y_{1i_1} \\
= & \sum_{l=1}^{n_1} a(i_1,l) y_{1l}^* + \sum_{l=1}^{n_2} a(i_1, n_1+l) y_{2l}^* - y_{1i_1} \\
= & \sum_{l \neq i_1} a(i_1,l) y_{1l} + \sum_{l \neq i_2} a(i_1,n_1+l) y_{2l} + a(i_1,i_1)\hat{f}_1^{\{i\}}(t_{1i_1}) + a(i_1,n_1+i_2)\hat{f}_2^{\{i\}}(t_{2i_2}) - y_{1i_1} \\
= & \sum_{l=1}^{n_1} a(i_1,l) y_{1l} + \sum_{l=1}^{n_2} a(i_1,n_1+l) y_{2l} - y_{1i_1} + a(i_1,i_1)(\hat{f}_1^{\{i\}}(t_{1i_1}) - y_{1i_1}) \\
& + a(i_1,n_1+i_2)(\hat{f}_2^{\{i\}}(t_{2i_2}) - y_{2i_2}) \\
= & \hat{f}_1(t_{1i_1}) - y_{1i_1} + a(i_1,i_1)(\hat{f}_1^{\{i\}}(t_{1i_1}) - y_{1i_1}) + a(i_1,n_1+i_2)(\hat{f}_2^{\{i\}}(t_{2i_2}) - y_{2i_2}).
\end{aligned}$$

Therefore

$$(1 - a(i_1,i_1))(\hat{f}_1^{\{i\}}(t_{1i_1}) - y_{1i_1}) - a(i_1,n_1+i_2)(\hat{f}_2^{\{i\}}(t_{2i_2}) - y_{2i_2}) = \hat{f}_1(t_{1i_1}) - y_{1i_1}.$$

The second equation in (13) can be derived similarly.

Proof of (20)

Let $A_k^a = (aT_kT_k^T + b_k\Sigma_k)(aT_kT_k^T + b_k\Sigma_k + \sigma_k^2 I)^{-1}$, $\tilde{\lambda}_k = \sigma_k^2/b_k$. It is known that $\lim_{a\to\infty} A_k^a = A_k$ and $\tilde{\boldsymbol{f}}_k = \lim_{a\to\infty} A_k^a \boldsymbol{y}_k = \lim_{a\to\infty} \mathrm{E}(\boldsymbol{F}_k|\boldsymbol{y}_k)$ (Wahba 1990). Therefore

$$\tilde{\boldsymbol{f}} - \boldsymbol{F} = \begin{pmatrix} \tilde{\boldsymbol{f}}_1 - \boldsymbol{F}_1 \\ \tilde{\boldsymbol{f}}_2 - \boldsymbol{F}_2 \end{pmatrix} = -\lim_{a\to\infty}[\begin{pmatrix} I - A_1^a & \\ & I - A_2^a \end{pmatrix}\begin{pmatrix} \boldsymbol{F}_1 \\ \boldsymbol{F}_2 \end{pmatrix} - \begin{pmatrix} A_1^a & \\ & A_2^a \end{pmatrix}\begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{pmatrix}]$$

with covariance

$$\begin{aligned}
& \lim_{a\to\infty}[\begin{pmatrix} I - A_1^a & \\ & I - A_2^a \end{pmatrix}\begin{pmatrix} aT_1T_1^T + b_1\Sigma_1 & \\ & aT_2T_2^T + b_2\Sigma_2 \end{pmatrix}\begin{pmatrix} I - A_1^a & \\ & I - A_2^a \end{pmatrix} \\
& \quad + \begin{pmatrix} A_1^a & \\ & A_2^a \end{pmatrix}\begin{pmatrix} \sigma_1^2 I & \sigma_1\sigma_2\rho I \\ \sigma_1\sigma_2\rho I & \sigma_2^2 I \end{pmatrix}\begin{pmatrix} A_1^a & \\ & A_2^a \end{pmatrix}] \\
= & \lim_{a\to\infty}[\begin{pmatrix} \sigma_1^2 A_1^a(I - A_1^a) & \\ & \sigma_2^2 A_2^a(I - A_2^a) \end{pmatrix} + \begin{pmatrix} \sigma_1^2 A_1^a A_1^a & \sigma_1\sigma_2\rho A_1^a A_2^a \\ \sigma_1\sigma_2\rho A_2^a A_1^a & \sigma_2^2 A_2^a A_2^a \end{pmatrix}] \\
= & \theta G
\end{aligned}$$

Proof of Theorem 1

Since $\boldsymbol{t}_1 = \boldsymbol{t}_2$, then $T_1 = T_2$, $\Sigma_1 = \Sigma_2$, $Q_{11} = Q_{21}$, $Q_{12} = Q_{22}$, $A_1 = A_2$ and

$$W^{-1} = \begin{pmatrix} rI & \rho I \\ \rho I & I/r \end{pmatrix}.$$

Let UDU^T be the eigenvalue eigenvector decomposition of $Q_{k2}^T \Sigma_k Q_{k2}$ and $\Gamma_k = Q_{k2} U$. U is orthogonal and D is diagonal with nonnegative elements. If Σ_k is full rank, the diagonal elements of D are positive. It is easy to check that $A_k = I - \Gamma_{k2} H_k^{-1} \Gamma_{k2}^T$, where $H_k = (I + D/\tilde{\lambda}_k)$. Note that $\tilde{\lambda}_1 = \sigma_1^2/b_1 = r\theta/b_1 = r\lambda_1$ and $\tilde{\lambda}_2 = \sigma_2^2/b_2 = \theta/(rb_2) = \lambda_2/r$. Let $\Gamma = \mathrm{diag}(\Gamma_1, \Gamma_2) = Q_2 \mathrm{diag}(U, U)$. It can be verified that $AW^{-1} = W^{-1} - E_1$ and $G = W^{-1} - E_2$, where

$$E_1 = W^{-1}\Gamma \begin{pmatrix} rH_1 & \rho I \\ \rho I & H_2/r \end{pmatrix}^{-1} \Gamma^T W^{-1}$$

and

$$E_2 = \Gamma \begin{pmatrix} rH_1^{-1} & \rho(H_1^{-1} + H_2^{-1} - H_1^{-1}H_2^{-1}) \\ \rho(H_1^{-1} + H_2^{-1} - H_1^{-1}H_2^{-1}) & H_2^{-1}/r \end{pmatrix} \Gamma^T.$$

Thus we need to show that $E_1 - E_2$ is positive semi-definite.

Lemma 2 *(a) $E_1 - E_2$ is positive semi-definite iff $\Gamma^T(E_1 - E_2)\Gamma$ is positive semi-definite. (b) If $\Gamma^T(E_1 - E_2)\Gamma$ is positive definite, then for any $\boldsymbol{x} \in R^n$, $\boldsymbol{x}^T(E_1 - E_2)\boldsymbol{x} = 0$ iff $\boldsymbol{x} \in$ span(T).*

[Proof] (a) $\Rightarrow$ Obvious.
$\Leftarrow$ Since $\Gamma = Q_2\mathrm{diag}(U, U)$ and U is orthogonal, then $Q_2^T(E_1 - E_2)Q_2$ is positive semi-definite. Since (Q_1, Q_2) is an orthogonal matrix, then any $\boldsymbol{x} \in R^n$ can be represented as $\boldsymbol{x} = Q_1\boldsymbol{\beta}_1 + Q_2\boldsymbol{\beta}_2$. Therefore

$$\begin{aligned} & \boldsymbol{x}^T(E_1 - E_2)\boldsymbol{x} \\ = \; & \boldsymbol{\beta}_1^T Q_1^T(E_1 - E_2)Q_1\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2^T Q_2^T(E_1 - E_2)Q_2\boldsymbol{\beta}_2 + 2\boldsymbol{\beta}_1^T Q_1^T(E_1 - E_2)Q_2\boldsymbol{\beta}_2 \\ = \; & \boldsymbol{\beta}_2^T Q_2^T(E_1 - E_2)Q_2\boldsymbol{\beta}_2 \\ \geq \; & 0, \end{aligned}$$

where the second equality holds because $Q_1^T E_1 = Q_1^T E_2 = 0$.
(b) $\Rightarrow$ From the inequality above, we must have $\boldsymbol{\beta}_2 = 0$, which implies that $\boldsymbol{x} = Q_1\boldsymbol{\beta}_1 \in$ span(T).
$\Leftarrow$ It is obvious since span$(T) =$ span(Q_1).
Using Lemma 2 and the fact that

$$\Gamma^T W^{-1}\Gamma = \begin{pmatrix} rI_{n_1-2} & \rho I_{n_1-2} \\ \rho I_{n_1-2} & I_{n_1-2}/r \end{pmatrix} \triangleq \tilde{W}^{-1},$$

we only need to show

$$\tilde{W}^{-1} \begin{pmatrix} rH_1 & \rho I \\ \rho I & H_2/r \end{pmatrix}^{-1} \tilde{W}^{-1} - \begin{pmatrix} rH_1^{-1} & \rho(H_1^{-1} + H_2^{-1} - H_1^{-1}H_2^{-1}) \\ \rho(H_1^{-1} + H_2^{-1} - H_1^{-1}H_2^{-1}) & H_2^{-1}/r \end{pmatrix}$$

is positive semi-definite. Notice that both H_1 and H_2 are diagonal, by switching columns and rows, three matrices above can be transformed into block diagonal forms. Therefore we only need to show all diagonal blocks are positive semi-definite. That is, we need to show

$$\begin{aligned} &\begin{pmatrix} r & \rho \\ \rho & 1/r \end{pmatrix} \begin{pmatrix} rh_1 & \rho \\ \rho & h_2/r \end{pmatrix}^{-1} \begin{pmatrix} r & \rho \\ \rho & 1/r \end{pmatrix} \\ - &\begin{pmatrix} r/h_1 & \rho(1/h_1 + 1/h_2 - 1/h_1h_2) \\ \rho(1/h_1 + 1/h_2 - 1/h_1h_2) & 1/rh_2 \end{pmatrix} \end{aligned} \tag{25}$$

is positive semi-definite, where $h_1 = 1 + d/\tilde{\lambda}_1$, $h_2 = 1 + d/\tilde{\lambda}_2$, and d is a diagonal element of D. The matrix (25) equals to

$$\frac{\rho^2 d^2}{(1 + d/\tilde{\lambda}_1)(1 + d/\tilde{\lambda}_2) - \rho^2} \begin{pmatrix} \frac{r}{\tilde{\lambda}_1^2 + \tilde{\lambda}_1 d} & -\frac{\rho}{(\tilde{\lambda}_1 + d)(\tilde{\lambda}_2 + d)} \\ -\frac{\rho}{(\tilde{\lambda}_1 + d)(\tilde{\lambda}_2 + d)} & \frac{1}{r(\tilde{\lambda}_2^2 + \tilde{\lambda}_2 d)} \end{pmatrix}.$$

Obviously that the above matrix is positive semi-definite. It is positive definite if $\rho \neq 0$, $\tilde{\lambda}_1 < \infty$, $\tilde{\lambda}_2 < \infty$ and $d > 0$.

References

Anderson, S. J., Jones, R. H. and Swanson, G. D. (1990). Smoothing polynomial splines for bivariate data, *SIAM Journal on Scientific and Statistical Computing* **11**: 749–766.

Box, G. E. P. and Draper, N. R. (1965). The Bayesian estimation of common parameters from several responses, *Biometrika* **52**: 355–365.

Dongarra, J. J., Moler, C. B., Bunch, J. R. and Stewart, G. W. (1979). *LINPACK: Users' Guide*, Siam, Philadelphia.

Fessler, J. A. (1991). Nonparametric fixed-interval smoothing with vector splines, *IEEE Transactions on Signal Processing* **39**: 852–859.

Gallant, A. R. (1975). Seemingly unrelated nonlinear regressions, *Journal of Econometrics* **3**: 35–50.

Gao, F. (1994). Fitting smoothing splines to data from multiple sources, *Communications in Statistics: Theory and Methods* **23**: 1665–1698.

Gu, C. (1989). RKPACK and its applications: Fitting smoothing spline models, *Proceedings of the Statistical Computing Section, ASA:* pp. 42–51.

Guo, W., Wang, Y. and Brown, M. B. (1999). A multiprocess state-space model for time series with pulse and changing baseline. To appear, *Journal of the American Statistical Association.*

Miller, J. J. and Wegman, E. J. (1987). Vector function estimation using splines, *Journal of Statistical Planning and Inference* **17**: 173–180.

Myers, D. E. (1991). Multidimensional multivariable smoothing, *Spatial Statistics and Imaging*, A. Possolo eds., Institute of Mathematical Statistics, Hayward, California.

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 B* **53**: 233–243.

Soo, Y. W. and Bates, D. M. (1996). Multiresponse spline regression, *Computational Statistics & Data Analysis* **22**: 619–631.

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline, *Journal of the Royal Statistical Society B* **45**: 133–150.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59.

Wahba, G. (1992). Multivariate functional and operator estimation, based on smoothing splines and reproducing kernels, *Nonlinear Modelling and Forecasting, SFI Studies in the Science of Complexity*, Vol. XII, 95-112, M. Casdagli and S. Eubank eds., Addison Wesley.

Wahba, G. and Wang, Y. (1993). Behavior near zero of the distribution of GCV smoothing parameter estimates for splines, *Statistics and Probability Letters* **25**: 105–111.

Wang, Y. (1998). Smoothing spline models with correlated random errors, *Journal of the American Statistical Association* **93**: 341–348.

Wang, Y. and Brown, M. B. (1996). A flexible model for human circadian rhythms, *Biometrics* **52**: 588–596.

Wegman, E. J. (1981). Vector spline and the estimation of filter functions, *Technometrics* **23**: 83–89.

Yee, T. and Wild, C. (1996). Vector generalized additive models, *Journal of the Royal Statistical Society B* **58**: 481–493.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association* **57**: 348–368.

Affiliation and Addresses

Yuedong Wang
Department of Statistics
and Applied Probability
University of California
Santa Barbara, CA 93106
yuedong@pstat.ucsb.edu
Fax: 805-893-2334

Wensheng Guo
Department of Biostatistics
and Epidemiology
University of Pennsylvania
Philadelphia, PA 19104
wguo@cceb.upenn.edu

Morton B. Brown
Department of Biostatistics
University of Michigan
Ann Arbor, MI 48109
mbb@sph.umich.edu