Nonparametric Regression Analysis of Longitudinal Data
Author(s): Joan G. Staniswalis and J. Jack Lee
Source: *Journal of the American Statistical Association*, Vol. 93, No. 444 (Dec., 1998), pp. 1403-1418
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2670055
Accessed: 22-04-2020 17:43 UTC

Nonparametric Regression Analysis of Longitudinal Data

Joan G. STANISWALIS and J. Jack LEE

Nonparametric methods are developed for estimating the dose effect when a response consists of correlated observations over time measured in a dose–response experiment. The methods can also be applied to data collected from a completely randomized design experiment. Methods are developed for the detection and description of the effects of dose, time, and their interaction. The methods allow for individual variation in the timing and number of observations. A generalization allowing baseline covariates to be incorporated is addressed. These results may be used in an exploratory fashion in the process of building a random-effects model for longitudinal data.

KEY WORDS: Growth curves; Kernel estimator; Principal component analysis; Repeated measures.

1. INTRODUCTION

This article proposes nonparametric methods for exploring and for testing hypotheses about the dose effect in a dose–response experiment when a response consists of correlated observations over time. Methods are developed for the nonparametric detection and description of interactions between dose and time. These methods do not require either independent replicate observations or measurements at prespecified fixed times. Two datasets are used to motivate and illustrate the methods developed in this article. Both of these datasets suffer missing data.

The first dataset was obtained from a Phase I study (Buzdar et al. 1994) to determine the safety of droloxifene in patients with advanced metastatic breast cancer. Five dose levels of droloxifene were selected for this longitudinal dose–response study. As is typical in Phase I studies, patients were not randomized to treatments. Rather, the lowest dose was administered to six patients once daily. After there was evidence that the current dose level was well tolerated, additional patients were entered at the next-highest dose level. Dose escalation was completed in 30 women with 6 women at each of the five preselected dose levels. There was no reduction of the assigned dose or termination of the therapy due to toxicity. Among other endpoints, follicle-stimulating hormone (FSH) levels were monitored at the time of entry to the study ($t = 0$) and thereafter at approximately 2-week intervals for a total of 6 weeks. One goal was to determine whether a clinically significant suppression of FSH occurs over time and whether the FSH suppression is dose dependent.

The second dataset was an in vivo tumor growth experiment (Heitjan, Manni, and Santen 1993). Tumors were established in mice by injecting breast cancer cells into two mammary fat pads per mouse. The mice with at least one established tumor were then size-matched and randomized to one of five dose groups of α-difluoromethylornithine (DFMO). There were 15 mice per dose group. The tumor volume was measured at 0, 3, 7, 10, 14, and 16 days posttreatment. When a mouse had more than one established tumor, the average log(volume) of the tumors was used. Of interest is an assessment of the effect of DFMO on tumor growth.

In both of these cases, the data are of the form $\mathbf{Y}_{ij}^t = (Y_{ij1}, \ldots, Y_{ijK(i,j)})$, for $j = 1, \ldots, n_i, i = 1, \ldots, m$, from a longitudinal dose–response experiment. Here Y_{ijk} is an observation taken on the jth individual assigned to dose d_i at time t_{ijk}, for $k = 1, \ldots, K(i,j)$. There are a total of $n = \sum_{i=1}^{m} n_i$ individuals in the study. The number of observations taken on the jth individual in the ith dose group is $K(i,j) \leq K$, where K denotes the maximum number of measurements over time and is not allowed to depend on n. For purposes of exposition, we will continue to refer to the explanatory variable d as a dose, although it could also be a vector of baseline covariates including dose or the treatment group indicator. This generalization is addressed later in the article; the baseline covariates could be a mixture of discrete and continuous variables.

One requirement of our methods is that as the sample size n increases, the pooled observation times become dense in the interval of observation. In the first dataset, we expect to have very few observations per individual. That is, $K(i,j)$ is bounded by a fixed number $K = 4$ not depending on n. The observation times t_{ijk} are random but clustered about the times 0, 14, 28, and 42 days. By letting the sample size n increase, we would expect the pooled measurement times to become dense in the interval of observation. In the second dataset, the time of measurement is a design variable controlled by the investigator. The number of observations $K(i,j)$ per animal is bounded by $K = 6$. If the sample size were increased, then for the pooled measurement times to become dense in the interval of observation, the experimenter would have to, for example, stagger the observation times within groups of animals.

Surveys of the parametric methods available for modelling the type of data considered here have been provided

Joan G. Staniswalis is Associate Professor, Department of Mathematical Sciences, University of Texas, El Paso, TX 79968-0514 (E-mail: joan@math.utep.edu). J. Jack Lee is Associate Professor, Department of Biomathematics, M.D. Anderson Cancer Center, University of Texas, Houston, TX 77030. This research was sponsored in part by National Institutes of Health grants S06GM08012, CA-16672, and CA-52051 and National Science Foundation grant DMS-9403582. The authors thank Diane Liu and Chi-Hong Tseng for programming assistance.

Journal of the American Statistical Association
December 1998, Vol. 93, No. 444, Theory and Methods

by Heitjan et al. (1993) and Laird, Donnelly, and Ware (1992). Among the methods surveyed are the growth curve model. As an illustration, Heitjan et al. (1993) applied the growth curve model to the data obtained from the DFMO study. There it is assumed that the log(tumor volume) increases linearly over time, but that the rate of growth may differ between treatment groups as a fixed effect and among individuals within a treatment group as a random effect. The authors described a popular approach in the medical research: fit a linear growth curve to the data collected from each animal, then conduct a multivariate analysis of variance (MANOVA) test for equality of the expected value of the regression coefficients between treatment groups.

Another model included in the survey is the random-effects model for longitudinal data (Laird and Ware 1982). This has become quite popular and by design can handle variation among individuals in both the number and timing of observations. Specifically, each individual response is modeled as $\mathbf{Y}_{ij} = \mathbf{X}_{ij}\alpha + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \mathbf{e}_{ij}$, where the $\mathbf{e}_{ij}$ is iid normal with mean 0 and covariance $\mathbf{R}_{ij}$. The $\mathbf{b}_{ij}$ is assumed to be normally distributed with mean 0 and covariance ϕ, independently of each other and of the $\mathbf{e}_{ij}$. Marginally, the $\mathbf{Y}_{ij}$ are independent normal with mean $\mathbf{X}_{ij}\alpha$ and covariance matrix $\mathbf{R}_{ij} + \mathbf{Z}_{ij}\phi\mathbf{Z}_{ij}^t$. An example of a random-effects model for the data from the droloxifene trial is

$$\begin{pmatrix} Y_{ij1} \\ \vdots \\ Y_{ijK(i,j)} \end{pmatrix} = \begin{pmatrix} 1 & d_i & t_{ij1} & d_i t_{ij1} \\ \vdots & & & \\ 1 & d_i & t_{ijK(i,j)} & d_i t_{ijK(i,j)} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_{12} \end{pmatrix} + \begin{pmatrix} 1 & t_{ij1} \\ & \vdots \\ 1 & t_{ijK(i,j)} \end{pmatrix} \begin{pmatrix} b_{0ij} \\ b_{1ij} \end{pmatrix} + e_{ij}.$$

The $\mathbf{X}_{ij}$ is the design matrix corresponding to the fixed effects of dose and time, which are linear in the parameters $\alpha_0, \ldots, \alpha_{12}$. Thus the average response for a treatment group is modeled with a common linear trend $\alpha_0 + \alpha_2 t$, a dose-by-time interaction $\alpha_{12} d \times t$, and a shift-by-dose $\alpha_1 d$. The matrix $\mathbf{Z}_{ij}$ is the design matrix linking the random effects b_{0ij} and b_{1ij} to Y_{ij}. The individual responses are allowed to vary about the group mean according to random shifts and a random linear trend. Estimates of b_{0ij} and b_{1ij} may be used to identify experimental units with unusual responses.

Notice that the growth curve model is a special case of the random-effects model that assumes that the linear space containing the fixed effects is a subspace of the linear space containing the random effects. That is, in the growth curve model it is assumed that the column space of $\mathbf{Z}_{ij}$ contains the column space of $\mathbf{X}_{ij}$. This modeling assumption limits the covariance structure of the individual random effects. Laird and Waire (1982, p. 972) pointed out that that this is undesirable because "we may want to fit a saturated model to the population growth curve and a very simple model to the individual deviations."

If true replicates are available within some of the dose groups, then a test for interaction between time and dose is performed by testing for lack of fit of a parametric model with additive effects. When true replicates are not available, tests for interactions are conducted by fitting nested parametric models with and without an interaction term. In this article we propose a test for interaction between time and dose with minimal modeling assumptions that does not require true replicates. The methods proposed herein may also be used in an exploratory fashion in building a parametric random-effects model for longitudinal data.

Nonparametric methods using kernel estimators have been considered for smoothing longitudinal or repeated-measures data (e.g., Altman 1990; Altman and Casella 1995; Fraiman and Meloche 1994; Hart and Wehrly 1986; and Müller 1988). All of these nonparametric approaches have the common feature that the unknown mean response curve over time is estimated by smoothing the raw data and time is the only explanatory variable. More recently, Zeger and Diggle (1994) studied a semiparametric model for longitudinal data in which the covariates were entered parametrically and only the time effect was entered nonparametrically. They extended to longitudinal data the backfitting algorithm of Hastie and Tibshirani (1990) for semiparametric regression. Zeger and Diggle (1994) provided additional references to nonparametric approaches to estimating the mean response over time given longitudinal data. A contribution of our article is that all explanatory variables are included into the data analysis with minimal modeling assumptions.

Many [Berkey and Kent (1983), Berkey, Laird, Valadign, and Gardner (1991), Besse and Ramsay (1986), Castro, Lawton, and Sylvestre (1986), Jones and Rice (1992), Ramsay and Dalzell (1991), Rice and Silverman (1991), and Silverman (1996)] have modeled longitudinal data nonparametrically by estimating the eigenfunctions corresponding to the covariance function of a random curve with continuous sample curves. A principal component analysis picks out the important features of a large collection of sample curves observed at fixed time points. The principal component analysis describes how individuals vary from the group mean over time. Berkey et al. (1991) used principal component scores for adolescent blood pressure measurements taken over time as covariates in a linear model to predict adult blood pressure. In all of these papers, the number and time of observations are either assumed to be the same for all the individuals in the study or interpolated to lie on a grid. A contribution of this article is that individual variation in the number and time of observation is allowed and interpolation is not needed in preparation for smoothing.

We adopt the following view. The sample is of the form (t_{ijk}, Y_{ijk}), where $Y_{ijk} = Y_{ij}(t_{ijk})$ and $Y_{ij}(t), t \in [0, T]$, is a stochastic process arising as the response of the jth replication at dose (or baseline covariate) d_i. The observation $Y_{ij}(t)$ is contaminated with an unobservable error process composed of white noise plus a smooth random function describing an individual's deviation from the group mean.

The model is

$$Y_{ij}(t) = \mu_i(t) + \eta_{ij}(t) + \varepsilon_{ij}(t), \tag{1}$$

where $\mu_i(t) = \mu(d,t)|_{d=d_i}$. The stochastic processes $\eta_{ij}(t)$ and $\varepsilon_{ij}(t)$ are independent with mean 0 and covariance functions $V(s,t)$ and $\sigma^2\delta(s,t)$, where $\delta(s,t) = 1$ if $s = t$ and 0 otherwise. Thus $Y_{ij}(t)$ has covariance function $C(s,t) = V(s,t) + \sigma^2\delta(s,t), 0 \le s,t \le T$. It is assumed that $\int_0^T \int_0^T |V(s,t)|^2\, dP(s)\, dP(t)$ exists. $P(t)$ denotes an absolutely continuous univariate cumulative distribution function obtained by taking the limit as $n \to \infty$ of the empirical distribution of the pooled visit times $\{t_{ijk}\}$. The underlying distribution of observation times within each dose group does not depend on the dose group or the baseline covariate. The covariance function $V(s,t)$ has continuous second-order derivatives. To show that this framework is immediately useful, in Section 2 we use the Karhunen–Loeve expansion for the process $\eta(t), t \in [0,T]$ to smooth the data obtained from each individual. A predictor of the smooth random curve $\eta_{ij}(t)$ for an individual is computed for the entire interval of observation. This will also lead to exploratory analysis that can be used to identify subjects with extreme deviations from the treatment group mean.

We pose two models for $\mu(d,t)$. The first model is highly structured to include a dose effect, a time effect, and a dose-by-time interaction as in the usual repeated-measures ANOVA model

$$\mu(d,t) = f_1(d)g_1(t) + f_2(d)g_2(t) + h(t), \tag{2}$$

where $g_1(t) = 1$. However, unlike the repeated-measures ANOVA model, the dose, time, and dose-by-time effects are modeled nonparametrically as smooth functions. Another difference between the repeated-measures ANOVA model and our approach is that we can handle individual variation in both the time of observation and the number of observations. The model given by (2) is a Tukey type of alternative to an additive model for $\mu(d,t)$. When replicate observations are not available in the usual repeated-measures ANOVA model, a test for the presence of an interaction between dose and time requires a parametric model for the interaction. An important contribution of this article is a test for an interaction between dose and time that does not require specification of a parametric form for the interaction term in the Tukey-type alternative. For the case when the random errors contaminating the observations are iid, Eubank, Hart, Simpson, and Stefanski (1995) analyzed a Tukey-type test of additivity that was proposed by Hastie and Tibshirani (1990).

The model given by (2) does not always hold. The second model studied for $\mu(d,t)$ is a generalization of (2),

$$\mu(d,t) = \sum_{i=1}^{\infty} f_i(d)g_i(t) + h(t), \tag{3}$$

where $\{g_i(t)\}$ form a complete orthonormal system in the $L_2(P)$ sense. This model for $\mu(d,t)$ always holds under some additional smoothness assumptions.

Our approach to modeling the group means $\mu(d,t)$ in this article is new in that it is based on the symmetric, nonnegative definite function defined as

$$A(s,t) = \frac{1}{n-1}\sum_{i=1}^{m} n_i[\mu_i(s) - \mu(s)][\mu_i(t) - \mu(t)],$$

where $\mu(t) = \sum_{i=1}^{m} \mu_i(t)n_i/n$. Nonparametric plug-in methods are used to estimate $A(s,t)$. Notice that $A(s,t)$ should store information about how group means $\mu_i(t)$ vary about the grand mean $\mu(t)$ in much the same way that the covariance function contains information about how individuals vary about their group mean.

For the model given by (2), the manner in which $A(s,t)$ is used for our purposes is introduced in Section 3. Testing the null hypotheses that $f_1 = 0$ or that $f_2 = 0$ is addressed in Section 4, where it is assumed for testing purposes only that the error term is a Gaussian process. The other model given by (3) is studied in Section 5. In Section 6 are some details about the construction of the estimators $\hat{A}(s,t), \hat{V}(s,t), \hat{\mu}(d,t)$, and $\hat{\sigma}^2$ that can be skipped by the reader who only wants an overview of our approach. The results of a computer simulation study for the methods based on (2) and (3) are presented in Sections 7 and 8. In Section 9 the two datasets are analyzed. The proofs of any assertions are relegated to the Appendix. A technical report with more detailed proofs is available from the first author (Staniswalis and Lee 1997). Without loss of generality, take $T = 1$ and $d \in [0,1]$ on standardized scales for time and dose.

2. SMOOTHING DATA FOR A SUBJECT

A predictor (Robinson 1991) for an individual subject's curve $\mu_i(t) + \eta_{ij}(t)$ may be constructed by applying a scatterplot smoother (Eubank 1988) to the data corresponding to that subject. However, a predictor with smaller mean squared error (MSE) may be obtained from the Karhunen–Loeve expansion for the process $\eta(t), t \in [0,1]$ with covariance function $V(s,t)$. This filters the random noise $\varepsilon_{ij}(t)$ from $Y_{ij}(t)$ as described herein.

Denote the eigenvalue–eigenfunction pairs of $V(s,t)$ with respect to $L_2(P)$ by $(\lambda_q, \Psi_q(\cdot))$, with $\lambda_1 \ge \lambda_2 \ge \cdots > 0$. Assume that the eigenfunctions are normalized. If V were known, then

$$\int_0^1 V(s,t)\Psi_i(t)\, dP(t) = \lambda_i\Psi_i(s)$$

and

$$\int_0^1 \Psi_j(t)\Psi_i(t)\, dP(t) = \delta(i,j).$$

An approximation of the best Q-dimensional linear model (Castro et al. 1988) for $Y_{ij}(t) - \varepsilon_{ij}(t) = \mu_i(t) + \eta_{ij}(t)$ is given by

$$\mu_i(t) + \sum_{q=1}^{Q} A_{ijq}\Psi_q(t),$$

where $A_{ijq} = \int_0^1 (Y_{ij}(t) - \mu_i(t))\Psi_q(t)\, dP(t)$. The approximation comes from the fact that $Y_{ij}(t)$ are contaminated

with the white noise $\varepsilon_{ij}(t)$. Note that the principal component scores A_{ijq} have mean 0.

Because $Y_{ij}(t)$ is not observed for all $t \in [0,1]$, rather than computing integrals as described earlier, we will be forced to approximate integrals with a quadrature rule. To this end, we now introduce some notation. Set $\mathbf{\Psi}_{ijq} = (\Psi_q(t_{ij1}), \ldots, \Psi_q(t_{ijK(i,j)}))^T$ and $\boldsymbol{\mu}_{ij} = (\mu_i(t_{ij1}), \ldots, \mu_i(t_{ijK(i,j)}))^T$. Define $a_{ijq} = (\mathbf{Y}_{ij} - \mu_{ij})^T \mathbf{W}_{ij} \mathbf{\Psi}_{ijq}, q = 1, \ldots, Q$, where the $\mathbf{W}_{ij}$ are quadrature weights further specified in the Appendix. The a_{ijq} still have zero expected value. The a_{ijq} are approximations of A_{ijq}. Now

$$\mu(d_i, t) + \sum_{q=1}^{Q} a_{ijq} \Psi_q(t) \tag{4}$$

is an approximation of the best Q-dimensional linear model for $Y_{ij}(t) - \varepsilon_{ij}(t) = \mu_i(t) + \eta_{ij}(t)$. This time the approximation comes from both using a quadrature rule rather than exactly computing the integrals, and from the fact that $Y_{ij}(t)$ is also contaminated with the white noise $\varepsilon_{ij}(t)$. Expression (4) provides a smoothed version of $Y_{ij}(t)$ with some of the roughness contributed by the white noise $\varepsilon_{ij}(t)$ removed.

In practice, $\mu(d_i, t)$ and Ψ_q in (4) must be estimated from the data. Let the estimators be denoted by " $\hat{}$ ". Because $\mu(d_i, t)$ will be estimated with scatterplot smoothers that are known to be biased estimators, $\hat{a}_{ijq}$ will not have zero expected value. Thus in (4), the estimated principal component scores should be centered by $\hat{a}_{i \cdot q} = \sum_j \hat{a}_{ijq}/n_i$. The predictor of a subject's curve is given by

$$\tilde{Y}_{ij}(t) = \hat{\mu}(d_i, t) + \sum_{q=1}^{Q} (\hat{a}_{ijq} - \hat{a}_{i \cdot q}) \hat{\Psi}_q(t).$$

This has been constructed from a truncated generalized Fourier series using estimators of the normalized eigenfunctions of $V(s,t)$ as elements of a basis. Figure 1 shows the $Y_{ij}(t)$ (solid line connecting the raw data points) and the smooth provided by $\tilde{Y}_{ij}(t)$ (dashed line) using $Q = 2$ for the droloxifene study. The advantage of using $\tilde{Y}_{ij}(t)$ over applying a scatterplot smoother to the subject curve is that the MSE of $\tilde{Y}_{ij}(t)$ will depend on the bandwidth through n rather than through K, the number of observation times per subject.

Extreme values of the centered principal component scores $\hat{a}_{ijq} - \hat{a}_{i \cdot q}$ can be used to identify outliers; that is, patients with extreme deviations from the respective group

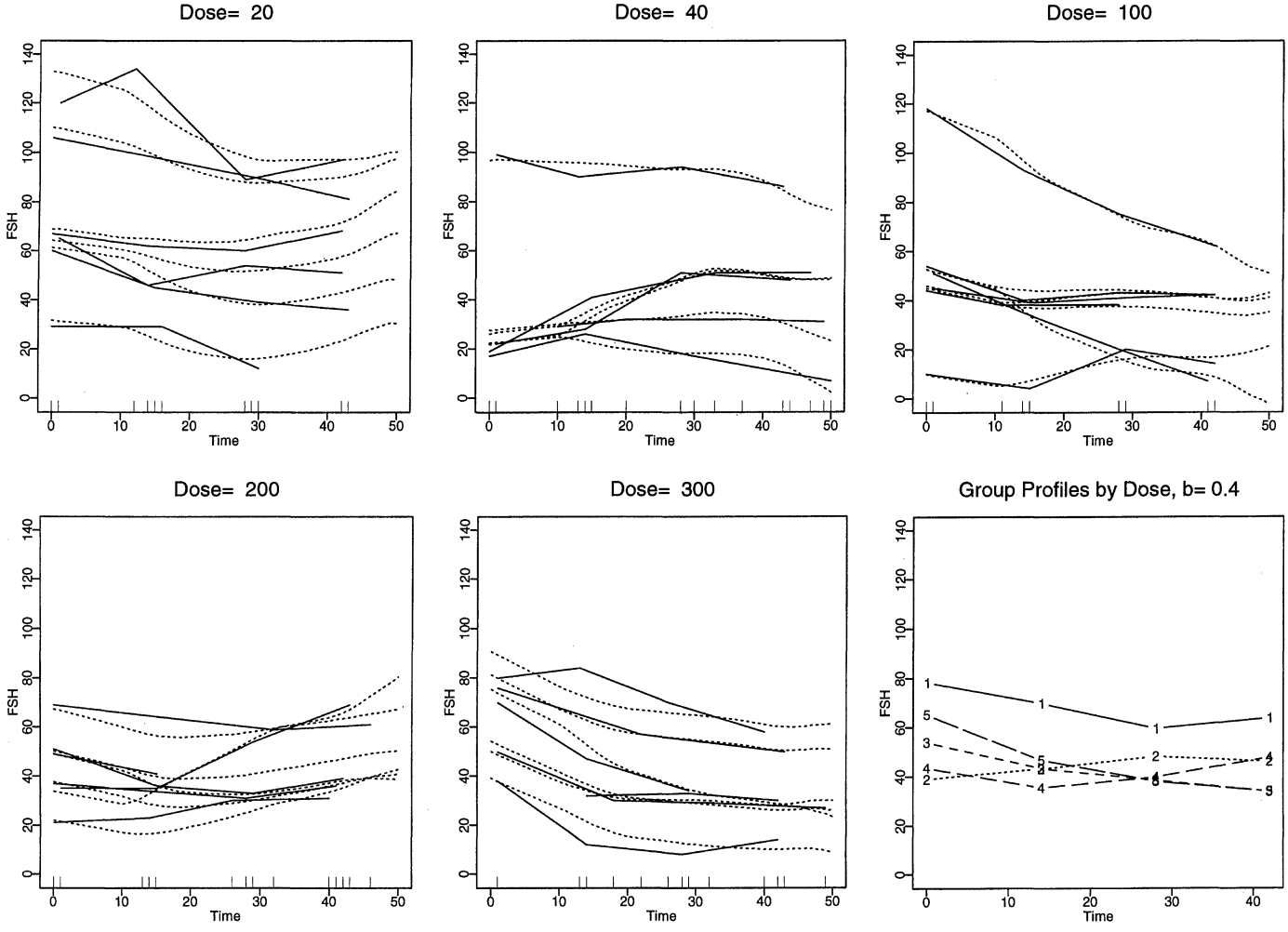

Figure 1. Patient Profiles of Droloxifene Phase I Trial: Raw Data (Connected by Solid Lines) and Smoothed With a Truncated Generalized Fourier Series (Dotted Line), Q = 2. Group profiles by dose are obtained by smoothing data aggregated by group. 1 ———, dose = 20; 2 · · ·, dose = 40, 3 - - -, dose = 100; 4 — — — dose = 200; 5 — — —, dose = 300.

mean. The centered principal component scores may also be used to identify associations between the response and underlying covariates (see Jones and Rice 1992).

3. A STRUCTURED NONPARAMETRIC REPEATED-MEASURES ANALYSIS OF VARIANCE MODEL

We first consider the structured model for the mean given by (2). If we have a completely randomized design, with no missing data or variation in the time of observation, then the data may be analyzed by the usual repeated-measures MANOVA. Our nonparametric structured model is like a smooth MANOVA, so it should be more efficient when the number of measurement times is large. It is assumed that $g_1(t) = 1$ but that h, g_2, f_1, and f_2 are unknown functions with two continuous derivatives on [0, 1] satisfying the following:

$$\int_0^1 g_i(t)g_j(t)\,dP(t) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \qquad \text{for} \quad i, j = 1, 2$$

and

$$\sum_{i=1}^m n_i f_1(d_i) = 0, \qquad \sum_{i=1}^m n_i f_2(d_i) = 0,$$

and

$$\sum_{i=1}^m n_i f_1(d_i) f_2(d_i) \geq 0.$$

Hence $\mu(t) = h(t)$.

This model for $\mu(d, t)$ includes a shift–dose effect, a time effect and a dose-by-time interaction as in the usual repeated-measures ANOVA model. The smooth function $f_2(d)$ measures the dose-by-time interaction. The function $f_1(d)$ provides a shift on the mean response for a given dose. Because $\mu(t) = h(t)$, the function $h(t)$ is the overall time effect on the mean response. These interpretations are shared by the usual repeated-measures ANOVA model. The example of a random-effects model for the DFMO data in Section 1 is in fact a structured model with a parametric choice for f_1, f_2, g_2, and h. Obviously, model (2) does not always hold. Yet it gives a class of models in which the effect on the overall outcome measured can be partitioned into the familiar dose effect, time effect, and dose-by-time interaction. The functional forms of these effects will be estimated from the data using nonparametric procedures.

Recall the definition of $A(s, t)$ given in Section 1. It can be shown that under (2),

$$A(s,t) = \tau_1 g_1(s)g_1(t) + \tau_{12}[g_1(t)g_2(s) + g_1(s)g_2(t)] + \tau_2 g_2(s)g_2(t),$$

where $\tau_r = (n-1)^{-1}\sum_{i=1}^m n_i f_r^2(d_i), r = 1, 2$, and $0 \leq \tau_{12} = (n-1)^{-1}\sum_{i=1}^m n_i f_1(d_i) f_2(d_i)$. When $\tau_1 = 0$, there is no dose-related shift in the average response over time. When $\tau_2 = 0$, there is no interaction between dose and time. It is easy to verify that

$$\int_0^1 \int_0^1 A(s,t)g_1(s)g_1(t)\,dP(s)\,dP(t) = \tau_1,$$

$$\int_0^1 [A(s,t) - \tau_1 g_1(s)g_1(t)]g_1(t)\,dP(t) = \tau_{12}g_2(s),$$

and

$$\int_0^1 \int_0^1 A(s,t)g_2(s)g_2(t)\,dP(s)\,dP(t) = \tau_2.$$

Let $P_n(t)$ denote the empirical cdf of the pooled patient visit times $t_{ijk}, k = 1, \ldots, K(i,j), j = 1, \ldots, n_i, i = 1, \ldots, m$. Then if we are given an estimator $\hat{A}(s,t)$ of $A(s,t)$, the estimators of the τ's and $g_2(\cdot)$ are taken as

$$\hat{\tau}_1 = \int_0^1 \int_0^1 \hat{A}(s,t)g_1(s)g_1(t)\,dP_n(s)\,dP_n(t),$$

$$\hat{g}_2(s) = c_1 \int_0^1 [\hat{A}(s,t) - \hat{\tau}_1 g_1(s)g_1(t)]g_1(t)\,dP_n(t) - c_2,$$

and

$$\hat{\tau}_2 = \int_0^1 \int_0^1 \hat{A}(s,t)\hat{g}_2(s)\hat{g}_2(t)\,dP_n(s)\,dP_n(t), \qquad (5)$$

where c_1 and c_2 are chosen so that $g_1(t)$ and $\hat{g}_2(t)$ are orthonormal in $L_2(P_n)$. If $\hat{A}(s,t)$ is a uniformly consistent estimator of $A(s,t)$, then $\hat{\tau}_1$ and $\hat{\tau}_2$ are consistent estimators of τ_1, and τ_2.

Estimation of $A(s,t)$ and $C(s,t)$ is addressed using ideas from nonparametric regression in Section 6. For now, just assume that $\hat{A}(s,t)$ and $\hat{C}(s,t)$ are uniformly consistent estimators of $A(s,t)$ and $C(s,t)$.

4. ESTIMATION AND TESTING OF THE DOSE EFFECT

Our goal here is to find projections of the data that reveal the shape of f_1 and f_2 under (2). The idea is similar in spirit to the parametric analysis reviewed by Heitjan et al. (1993), where a linear growth curve is fit to the log(tumor volume) from each animal and then a MANOVA test for equality of the expected value of regression coefficients between treatment groups is conducted. Rather than obtain regression coefficients by fitting, say, a linear growth model to the data, we project the data onto smooth functions.

For example, consider the result of projecting $Y_{ij}(t) - h(t)$ in the direction of the smooth function $g_1(t)$:

$$\int_0^1 (Y_{ij}(t) - h(t))g_1(t)\,dP(t) = f_1(d_i) + \int_0^1 \nu_{ij}(t)g_1(t)\,dP(t),$$

where $\nu_{ij}(t) = \eta_{ij}(t) + \varepsilon_{ij}(t)$. A scatterplot of the projection $\int_0^1 (Y_{ij}(t) - h(t))g_1(t)\,dP(t)$ versus dose reveals the shape of f_1 contaminated with iid errors. Similarly, a scatterplot of the projection

$$\int_0^1 (Y_{ij}(t) - h(t))g_2(t)\,dP(t) = f_2(d_i) + \int_0^1 \nu_{ij}(t)g_2(t)\,dP(t),$$

versus dose reveals the shape of f_2 contaminated with iid errors. If $f_1(d)$ and $f_2(d)$ are smooth functions of dose, then any scatterplot smoother may be applied to these iid data to obtain a consistent estimator of this dose effect (Eubank 1988). Under the assumption that $\varepsilon_{ij}(t)$ and $\eta_{ij}(t)$ are Gaussian processes, the hypotheses H_0: $\tau_1 = 0$ and H_0: $\tau_2 = 0$ can be tested by using a one-way ANOVA of the projections. The ANOVA test statistics for a completely randomized design based on this projected data are F distributed with $m-1$ and $n-m$ df under the null hypothesis. When dose is not a factor in a completely randomized design, the one-way ANOVA is not appropriate. If, for example, dose is a continuous explanatory variable, then other tests of the null hypotheses $\tau_r = 0$ are available in the literature, $r = 1, 2$. These methods test the hypotheses of no predictor effect in regression. Eubank and Hart (1993) proposed some new methods based on smoothers that are especially useful when the spectrum of the alternative hypotheses has high frequency components. Buckley (1991) proposed a cusum test, derived from work of Cox, Koh, Wahba, and Yandell (1988), that is best for detecting smooth departures from the null hypotheses $f_r = 0, r = 1, 2$.

Let $\tilde{W}_{ij}^{(1)} = \int_0^1 (\tilde{Y}_{ij}(t) - \hat{h}(t))g_1(t)\,dP_n(t)$ and $\tilde{W}_{ij}^{(2)} = \int_0^1 (\tilde{Y}_{ij}(t) - \hat{h}(t))\hat{g}_2(t)\,dP_n(t)$. As argued earlier, a scatterplot of $\tilde{W}_{ij}^{(1)}$ and $\tilde{W}_{ij}^{(2)}$ versus dose should reveal the shapes of f_1 and f_2. Now, because $h(t), g_2(t)$, and $P(t)$ must be estimated from the data, $W_{ij}^{(r)}(j = 1, \ldots n_i, i = 1, \ldots, m)$ are neither iid nor Gaussian for $r = 1, 2$. However, estimation of f_r using nonparametric smoothers and testing of H_0: $\tau_r = 0$ could proceed, assuming that these random variables are approximately iid and normal. Note that we do not use our estimates of $\hat{f}_1$ and $\hat{f}_2$ to estimate the value of $\tau_r, r = 1, 2$, but rather use (5).

The fact that $g_2(t)$ is estimated from the data introduces an additional complication in the testing of H_0: $\tau_2 = 0$. The size of the test is not maintained if the p value obtained from the one-way ANOVA is compared directly to the significance level α. We explain this difficulty and propose a Bonferonni correction by considering the case where $\mu(d, t) = 0$. The smooth random patient effects $\eta_{ij}(t)$ are linear combinations of $\Psi_1(t), \Psi_2(t), \ldots$, the eigenfunctions of $V(s, t)$. The coefficients in the linear combination are given by the principal component scores A_{ijq}. By chance alone, $A_{i\cdot q}$, the sample mean of the principal component scores along the principal direction $\Psi_q(t)$, could be significantly different between the dose groups. If by chance alone the $A_{i\cdot q}$ are significantly different between dose groups, then the estimate of $\hat{g}_2(t)$ will follow the smooth shape of $\Psi_q(t)$. That is, the contribution $A_{i\cdot q}\Psi_q(t)$ will be incorrectly interpreted as coming from the dose effect $f_2(d_i)g_2(t)$. This in turn will result in a significant one-way ANOVA for H_0: $\tau_2 = 0$ due to chance alone. Thus in general, if a significance level of α is used to test the hypothesis H_0: $\tau_2 = 0$, the ANOVA test statistic will have a size bounded by $Q'\alpha$. Because $g_2(t)$ must be orthogonal to $g_1(t), Q'$ is the number of directions $\Psi_1(t), \ldots, \Psi_Q(t)$ that are orthogonal to $g_1(t)$. Thus, asymptotically, the size of the test can be maintained below α by using the significance level α/Q' to test the hypothesis.

4.1 An Optimal Direction

In the previous section we showed that information about f_1 was obtained by projecting along the curve g_1. Similarly, information about f_2 was obtained by projecting along the curve g_2. In this section we investigate the choice of an optimal projection for minimizing the variance of the projections. By using an optimal projection, we increase the power of tests of hypotheses concerning f_1 and f_2.

Let $v(t)$ be any function in $L_2(P)$. Let $(r, r') = (1, 2)$ or (2, 1). Observe that if

$$\int_0^1 v(t)g_{r'}(t)\,dP(t) = 0$$

and

$$\int_0^1 v(t)g_r(t)\,dP(t) \neq 0, \tag{6}$$

then

$$\frac{\int_0^1 (Y_{ij}(t) - h(t))v(t)\,dP(t)}{\int_0^1 v(t)g_r(t)\,dP(t)} = f_r(d_i) + \frac{\int_0^1 v(t)\nu_{ij}(t)\,dP(t)}{\int_0^1 v(t)g_r(t)\,dP(t)}.$$

As before, our multivariate problem can be reduced to a univariate problem in which a scatterplot of the projections

$$\frac{\int_0^1 (Y_{ij}(t) - h(t))v(t)\,dP(t)}{\int_0^1 v(t)g_r(t)\,dP(t)}$$

versus dose d_i will reveal the shape of f_r. Define

$$\tilde{W}_{ij}(r, v) = \frac{\int_0^1 (\tilde{Y}_{ij}(t) - \hat{\mu}(t))v(t)\,dP_n(t)}{\int_0^1 v(t)\hat{g}_r(t)\,dP_n(t)}. \tag{7}$$

It is of interest to find a direction $v(\cdot)$ for projection of the $\tilde{Y}_{ij}$ that has the property of minimizing, with respect to the function $v \in L_2(P)$, the variance of the projection $\tilde{W}_{ij}(r, v)$ for $i = 1, \ldots, m$. Let $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dP(t)$. Assuming that $\int_0^T \int_0^T |V(s, t)|^2\,dP(s)\,dP(t)$ is finite, $V(s, t) = \lim_{N\to\infty} \sum_{i=1}^N \lambda_i \Psi_i(s)\Psi_i(t)$ uniformly on $[0, 1]^2$ (Griffel 1981, an adaptation of thm. 8.54, p. 233; Heuser 1982, p. 225, prop. 70.1, p. 272, prop. 71.1, p. 278). Treating Q as fixed in the definition of Y_{ij}, the variance of $\tilde{W}_{ij}(r, v)$ conditional on the visit times is approximately

$$\frac{\int_0^1 \int_0^1 v(s)v(t)[\sum_{q=1}^Q (\lambda_q + \sigma^2 \Psi_{ijq}^t \mathbf{W}_{ij}^2 \Psi_{ijq}) \times \Psi_q(s)\Psi_q(t)]\,dP(s)\,dP(t)}{(\int_0^1 v(t)g_r(t)\,dP(t))^2}. \tag{8}$$

Assuming that the patient visit times are clustered about $T_1, \ldots, T_K$, we can approximate (8) with

$$\frac{\int_0^1 \int_0^1 v(s)v(t)[\sum_{q=1}^Q (\lambda_q + \sigma_q^2)\Psi_q(s)\Psi_q(t)]\,dP(s)\,dP(t)}{(\int_0^1 v(t)g_r(t)\,dP(t))^2}, \tag{9}$$

where $\sigma_q^2 = \sigma^2 \sum_{i=1}^K \Psi_q^2(T_i)w_i^2$ and $\mathbf{W} = \text{diagonal}(w_1, \ldots, w_K)$ contains the quadrature weights needed to make

$\Psi_1(t), \ldots, \Psi_Q(t)$ orthonormal on the grid $\{T_1, \ldots, T_K\}$ as described in the Appendix. It is of interest to solve for $v_r^*(t) \in L_2(P)$ that minimizes (9) with respect to $v \in L_2(P)$ subject to (6) to minimize the variance in the scatterplot.

In our search for an optimal direction for use in the projection of the data, we make the following assumption.

Assumption A. For the structured nonparametric model, $g_r(t) = \sum_{i=1}^{Q} a_{ri}\Psi_i(t)$, where $a_{ri} = \langle g_r, \Psi_i \rangle$ for $r = 1, 2$.

This assumption is satisfied by, for example, the growth curve models reviewed in Section 1. There it was pointed out that in the parametric case, Laird and Ware (1982) found this assumption to be undesirable. This is not a valid criticism for the structured nonparametric model under consideration. For the analogous parametric random-effects model with $g_2(t)$ known, we are only imposing the reasonable assumption that the rank of the matrix $\mathbf{Z}$ is at least two. The validity of assumption A must be checked with the data before using the optimal projection described here. The following theorem provides the optimal direction obtained under assumption A. This direction is suboptimal when assumption A is violated. Proofs of the theorem and the following corollary are given in the Appendix.

Theorem 1. The optimal direction can be found by minimizing (9) subject to (6), under assumption A. This optimal direction is any nonzero scalar multiple of $v_r^*(t) = \sum_{q=1}^{Q} v_{rq}^* \Psi_q(t)$, where

$$v_{rq}^* = (\lambda_q + \sigma_q^2)^{-1}(V_{1r'}a_{rq} - V_2 a_{r'q}),$$

$$V_{1r'} = \sum_{q=1}^{Q} (\lambda_q + \sigma_q^2)^{-1} a_{r'q}^2,$$

and

$$V_2 = \sum_{q=1}^{Q} (\lambda_q + \sigma_q^2)^{-1} a_{rq}a_{r'q}. \tag{10}$$

A sufficient condition for the direction of the optimal projection $v_r^*(t)$ to coincide with the nonparametric functions $g_r(t)$ is provided next.

Corollary 1. Let $\eta(t) = U_1 g_1(t) + U_2 g_2(t)$ in (1), where $(U_1, U_2)^t$ follows a bivariate distribution with mean 0 and positive definite covariance matrix $\mathbf{\Phi}$. Then the optimal directions are $v_r^*(t) = g_r(t)$, for $r = 1, 2$.

4.2 Implementation of the Optimal Direction

First, estimate in $L_2(P)$ the eigenvalues and eigenfunctions of $V(s,t)$ on the grid $\{t_{ijk}\}$ using $\hat{V}(s,t)$, given in Section 5. Let $s_1 \le s_2 \le \ldots$ denote the sorted values of t_{ijk}. Now create a matrix $\hat{\mathbf{V}}$ with (i,j)th entry given by $\hat{V}(s_i, s_j)$. Find the eigenvalues and eigenvectors of the matrix $\hat{\mathbf{V}}$. The eigenvalues of $\hat{\mathbf{V}}$ provide an estimate of the eigenvalues of $V(s,t)$. Further, the eigenvectors of $\hat{\mathbf{V}}$ provide an estimate of the eigenfunctions $\Psi(t)$ on the grid $s_1 \le s_2 \le \ldots$. Denote the estimates of the eigenfunctions on the grid by $\hat{\Psi}_q(t)$ and the corresponding eigenvalues by $\hat{\lambda}_q$ (for additional details, see Castro et al. 1986). Assumption A implies that we must have $Q > 1$ (i.e., at least two nonzero eigenvalues) to compute the optimal direction as proposed in the previous section. Further, the optimal direction is valid only if assumption A is satisfied, that is, if $g_1(t), g_2(t) \in \text{span}\{\Psi_1(t), \ldots, \Psi_Q(t)\}$. Assumption A can be checked by regressing g_1 and $\hat{g}_2$ on $\hat{\Psi}_1(t), \ldots, \hat{\Psi}_Q(t)$ without an intercept. Assumption A is supported by the data if the predicted values for g_1 and $\hat{g}_2$ are approximately equal to g_1 and $\hat{g}_2$.

Under assumption A, an estimate $\hat{v}_r^*(t) = \sum_{q=1}^{Q} \hat{v}_{rq}^* \hat{\Psi}_q(t)$ of the optimal direction $v_r^*(t)$ may be obtained, where the parameters a in (10) are estimated by $\hat{a}_{rq} = \int_0^1 \hat{g}_r(t)\hat{\Psi}_q(t)\, dP_n(t)$. Let $\tilde{W}_{ij}^{*(r)}$ denote $\tilde{W}_{ij}(r,v)$ constructed using $v(t) = \hat{v}_r^*(t)$ in (7). An analysis may now be carried out using $\tilde{W}_{ij}^{*(r)}$ in the same manner that the $\tilde{W}_{ij}^{(r)}$ were used earlier. The Bonferroni-type adjustment as described earlier in Section 4 for testing H_0: $\tau_2 = 0$ is needed here as well when the optimal projection $\hat{v}_2^*(t)$ is used.

Approximate expressions for the variance of $\tilde{W}_{ij}^{(r)}$ and for the variance of the optimal projections are provided in the proof of Theorem 1 that may be used for power calculations associated with tests of hypotheses concerning τ_1 and τ_2. These expressions for the variance of the projections are approximate when the visit times vary among experimental units, because the latter variation is not taken into account. This is investigated in Section 7.

Under assumption A, an approximate expression for the variance of $\tilde{W}_{ij}^{(r)}$ is given by

$$\sum_{q=1}^{Q} (\lambda_q + \sigma_q^2)a_{rq}^2 \quad \text{for} \quad r = 1, 2. \tag{11}$$

Likewise, under assumption A, an approximate expression for the variance of the optimal projections $\tilde{W}_{ij}^{*(r)}$ is given by

$$\frac{\sum_{q=1}^{Q}(\lambda_q + \sigma_q^2){v_{rq}^*}^2}{(\sum_{q=1}^{Q} v_{rq}^* a_{rq})^2} \quad \text{for} \quad r = 1, 2. \tag{12}$$

5. THE GENERAL NONPARAMETRIC MODEL

The exploratory analysis of the data for acquiring information about $\mu(d,t)$ is described for a nonparametric model that subsumes (2). The model is

$$\mu(d,t) = \sum_{i=1}^{\infty} f_i(d)g_i(t) + h(t),$$

where $\{g_i\}_{i=1}^{\infty}$ form a complete orthonormal system for the space of square-integrable functions with the usual inner product on $L_2(P)$, g_i are estimated from the data, $\mu(d,t)$ has continuous second-order partial derivatives, and $\sum_{j=1}^{m} n_j f_i(d_j) = 0$ for all i. Our assumptions on $\mu(d,t)$

guarantee that $f_i(d)$ are smooth functions. Under (3),

$$A(s,t) = \sum_{k=1}^{\infty} \tau_k g_k(s) g_k(t) + \sum_{p=1}^{\infty} \sum_{r<p} \tau_{rp}(g_r(s)g_p(t) + g_r(t)g_p(s)).$$

Here $\tau_k = (n-1)^{-1} \sum_{i=1}^{m} n_i f_k^2(d_i)$ and $\tau_{rs} = (n-1)^{-1} \sum_{i=1}^{m} n_i f_r(d_i) f_s(d_i)$. Note that $\int_0^1 \int_0^1 A(s,t) g_k(s) g_k(t)\, dP(s)\, dP(t) = \tau_k$. Without loss of generality, it is assumed that the $\{g_k\}_{k=1}^{\infty}$ are ordered in such a way that $\{\tau_k\}$ is a decreasing sequence in k. Choose $\{g_i(t)\}_{i=1}^{\infty}$ to be a complete orthonormal basis for $L_2(P)$ with a maximum value for $\{\tau_i\}_{i=1}^{\infty}$. One way of choosing $\{\hat{g}_i\}_{i=1}^{T}$ is to take the eigenfunctions corresponding to the T largest positive eigenvalues $\{\hat{\tau}_i\}_{i=1}^{T}$ of $\hat{A}(s,t)$. This works because the eigenfunctions of $A(s,t)$ can be shown to maximize the integral equation $\int_0^1 \int_0^1 A(s,t) g_k(s) g_k(t)\, dP(s)\, dP(t)$ with respect to g_k subject to the orthonormality constraint. The τ's are estimated with

$$\hat{\tau}_k = \int_0^1 \int_0^1 \hat{A}(s,t) \hat{g}_k(s) \hat{g}_k(t)\, dP_n(s)\, dP_n(t). \tag{13}$$

To estimate the curve f_k, we project the smoothed data along the direction $\hat{g}_k$. Set $\tilde{W}_{ij}^{(k)}$ equal to $\tilde{W}_{ij}(r,v)$ in (7) with $v(t) = \hat{g}_k(t)$, and $r = k$. It follows that a scatterplot of $\tilde{W}_{ij}^{(k)}$ versus d_i will reveal the shape of f_k contaminated with noise. If $n_i > 1$ for $i = 1, \ldots, m$, then the null hypotheses H_0: $\tau_k = 0$ can be tested using a one-way ANOVA of the data $\tilde{W}_{ij}^{(k)}$. The one-way ANOVA test statistic is only approximately F distributed with $m-1$ and $n-m$ df. The variance of the projections is approximated by (8).

Recall that the main random patient effects occur in the directions $\Psi_1(t), \ldots, \Psi_Q(t)$ given by the first Q eigenfunctions of $V(s,t)$. By chance alone, the mean of the random effects in any one or more of the directions $\Psi_1(t), \ldots, \Psi_Q(t)$ could be significantly different between the dose groups. Thus in general, if a significance level of α is used to test the hypotheses H_0: $\tau_k = 0$, because $\hat{g}_k(t)$ is estimated from the data, the ANOVA test statistic will have a size bounded by $Q\alpha$. Asymptotically, the size of the test can be maintained by using the significance level α/Q.

Remark. The data could also be modeled semiparametrically with (3) assuming that only the functions of dose $\{f_i\}_{i=1}^{\infty}$ are unknown and that the $\{g_i\}_{i=1}^{\infty}$ is a set of completely specified functions that forms a complete orthonormal system in $L_2(P)$. The tests of hypotheses on the τ's proceeds with $\tilde{W}_{ij}^{(k)}$, except that $\{g_i\}_{i=1}^{\infty}$ does not need to be estimated from the data. Because g_k is not estimated from the data, a Bonferroni-type adjustment is not needed for tests of hypotheses.

6. ESTIMATORS OF $A(s, t)$ AND $C(s, t)$

Estimation of $A(s,t)$ and $C(s,t)$ are addressed using ideas from nonparametric regression. The kernel estimator is applied, although local polynomials or smoothing splines could be used. Let b and h denote bandwidths that determine the degree of smoothing imposed on the time and dose scales. The fact that these bandwidths depend on n has been suppressed from the notation. Let $w(s, \cdot; b)$ be a boundary corrected kernel (Rice 1984) for estimating the value of a nonparametric regression function at the point s using the bandwidth b. Let

$$\hat{\mu}_i(s) = \sum_{j=1}^{n_i} \sum_{k=1}^{K(i,j)} w(s, t_{ijk}; b) Y_{ijk} \div \left(\sum_{j=1}^{n_i} \sum_{k=1}^{K(i,j)} w(s, t_{ijk}; b) \right),$$

$$\hat{\mu}(d,s) = \sum_{l=1}^{m} w(d, d_l; h) \left(\sum_{u=1}^{n_l} \sum_{v=1}^{K(l,u)} w(s, t_{luv}; b) \right) \hat{\mu}_l(s) \div \left(\sum_{l=1}^{m} \sum_{u=1}^{n_l} \sum_{v=1}^{K(l,u)} w(d, d_l; h) w(s, t_{luv}; b) \right),$$

$$\hat{\mu}(s) = \sum_{i=1}^{m} \hat{\mu}(d_i, s) n_i / n,$$

$$\hat{A}(s,t) = \frac{1}{n-1} \sum_{i=1}^{m} n_i [\hat{\mu}(d_i, s) - \hat{\mu}(s)][\hat{\mu}(d_i, t) - \hat{\mu}(t)],$$

$$\tilde{V}(s,t) = \left[\sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{k,k'=1; k \neq k'}^{K(i,j)} w(s, t_{ijk}; b) w(t, t_{ijk'}; b) \times (Y_{ijk} - \hat{\mu}(d_i, t_{ijk}))(Y_{ijk'} - \hat{\mu}(d_i, t_{ijk'})) \div c_{ij}(t_{ijk}, t_{ijk'}) \right] \div \left[\sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{k,k'=1; k \neq k'}^{K(i,j)} w(s, t_{ijk}; b) w(t, t_{ijk'}; b) \right],$$

and

$$\hat{v}(s) = \left[\sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{k=1}^{K(i,j)} w(s, t_{ijk}; b) \times (Y_{ijk} - \hat{\mu}(d_i, t_{ijk}))^2 / c_{ij}(t_{ijk}, t_{ijk}) \right] \div \left[\sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{k=1}^{K(i,j)} w(s, t_{ijk}; b) \right] - \tilde{V}(s,s),$$

where

$$c_{ij}(s,t) = \begin{cases} 1 - \dfrac{\sum_r w(t,t_{ijr};b)}{\sum_q \sum_r w(t,t_{iqr};b)} - \dfrac{\sum_r w(s,t_{ijr};b)}{\sum_q \sum_r w(s,t_{iqr};b)} \\ \quad + \dfrac{\sum_q \sum_r \sum_h w(s,t_{iqr};b) w(t,t_{iqh};b)}{[\sum_q \sum_r w(s,t_{iqr};b)][\sum_g \sum_h w(t,t_{igh};b)]} \\ \qquad\qquad \text{if } n_i > 1 \\ 1 \qquad\qquad \text{otherwise.} \end{cases}$$

Finally, set

$$\hat{V}(s,t) = \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{k'=1}^{K(i,j)} \sum_{k=1}^{K(i,j)} w(s,t_{ijk};b) w(t,t_{ijk'};b) \times [(Y_{ijk} - \hat{\mu}(d_i,t_{ijk}))(Y_{ijk'} - \hat{\mu}(d_i,t_{ijk'})) \div c_{ij}(t_{ijk},t_{ijk'}) - \hat{\sigma}^2 \delta(t_{ijk},t_{ijk'})] \right\} \div \left[\sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{k'=1}^{K(i,j)} \sum_{k=1}^{K(i,j)} w(s,t_{ijk};b) w(t,t_{ijk'};b) \right]$$

and $\hat{C}(s,\ t) = \hat{\mathbf{V}}(s,\ t) + \hat{\sigma}^2 \delta(s,\ t)$, where $\hat{\sigma}^2 = \int_a^b \hat{v}(t)\ dP_n(t)$ and $a, b \in [0,1]$ are chosen to satisfy $\int_0^a dP_n(t) = \int_b^1 dP_n(t) = .25$. The term $c_{ij}(s,t)$ that appears inside the sum of the definition of the estimator of $V(s,t)$ may seem odd until one looks at the "internal" kernel-type estimators of Jones, Davies, and Park (1994).

The dependence of the latter quantities on the smoothing parameters is suppressed in this notation. The following additional assumptions are needed to establish that $\hat{A}(s,t)$ and $\hat{C}(s,t)$ are uniformly consistent for $A(s,t)$ and $C(s,t)$:

a. The kernel $w(s,t;b)$ has compact support and is uniformly bounded for $s,t \in [0,1]$ and $\mathbf{b} > 0$. The kernel is continuously differentiable in s and t.
b. Both $\hat{\mu}(s)$ and $\hat{\mu}(d,s)$ are uniformly consistent estimators of $\mu(s)$ and $\mu(d,s)$.
c. The distribution $P(t)$ is bounded away from 0 and continuously differentiable on its support.
d. For the situation where d is a baseline covariate measured on each individual rather than a design variable, the marginal distribution of d is bounded away from 0 and continuously differentiable on [0, 1]. The explanatory variable d could be taken to be a vector of baseline covariates (see Staniswalis and Messer 1996 and Staniswalis, Messer, and Finston 1994 for a discussion of kernel estimators for multiple regression.)

Cheng and Lin (1981) and Nadaraya (1989, p. 122) have discussed sufficient conditions on b and h to ensure uniform consistency of a kernel estimator applied to iid data. To apply those results here, we assume that there exists a sequence $a_n > 0$ such that $\inf_{ijk} |t_{ijk} - t_{ijk-1}| > a_n$ and $\liminf a_n > 0$. This assumption is needed to ensure that for very large $n, 2b < a_n$ can be chosen so that we are not smoothing over correlated observations over time. Fraiman and Meloche (1994) described situations where smoothing correlated observations over time is worse than no smoothing over time; that is, smoothing the averaged sample paths does not reduce the MSE. This can occur when, for example, the bound K on the number of observations per subject is allowed to grow with n too fast. We avoid this entirely by assuming that K does not depend on n, although we conjecture that this condition can be relaxed.

The following theorem establishes consistent estimators of $C(s,t), V(s,t)$, and $A(s,t)$ on which our procedures are based.

Theorem 2. Assume that we are working under the model given by (1). If assumptions a–d hold, then $\hat{C}(s,t), \hat{V}(s,t)$, and $\hat{A}(s,t)$ are uniformly consistent estimators of $C(s,t), V(s,t)$, and $A(s,t)$, as $n \to \infty$.

7. SIMULATIONS FOR THE NONPARAMETRIC REPEATED-MEASURES ANALYSIS OF VARIANCE MODEL

Here we report the results of a small simulation study conducted to investigate the proposed method for the model given by (1) and (2). The simulations were written in S-PLUS and executed on a Dec Alpha workstation. The S-PLUS pseudo-random number generators rnorm and runif were used to generate the data. The purpose of the simulations is to evaluate the size and power of the tests of hypotheses concerning τ_1 and τ_2 that use the projections and the F-distribution approximation.

The value of Q used in the simulation study was not chosen from the data. In the data analysis of Section 9 we demonstrate the considerations necessary for proper selection of Q. These considerations are difficult to implement in a simulation study. We used the quartic kernel $w(v) = (15/16)(1 - v^2)^2 I_{[-1,1]}(v)$ for the smoothing with the boundary modification of Rice (1984).

First, we set $d = 0, 1 (m = 2), n_1 = n_2 = 15$, and $K(i,j) = K = 7$ for all of the 100 simulated datasets generated for this part of the Monte Carlo study. The following assignments specify the visit times and $\mathbf{Y}_{ij}$:

- $h(t) = 0, f_1(d) = a_1(1 - 2d)$, and $f_2(d) = a_2(1 - 2d)$:
- The visit times were generated as random in the first part of the simulation study. The entire simulation was then rerun using fixed visit times. For the random visit times, t_{ijk} are independent and uniformly distributed on $(T_k - \frac{1}{14}, T_k + \frac{1}{14})$ for $k = 1, \ldots, 7$ with $T_1 = \frac{1}{14}, T_2 = \frac{3}{14}, \ldots, T_7 = \frac{13}{14}$. For the fixed visit times, $t_{ijk} = T_k$ for all i, j, k.
- The eigenfunctions of $V(s,t)$ are $\Psi_1(t) = 1, \Psi_2(t) = 2\sqrt{3}(t - \frac{1}{2})$, and $\Psi_3(t) = 6\sqrt{5}(t^2 - t + \frac{1}{6})$ with corresponding eigenvalues $\lambda_1 = 7, \lambda_2 = 4$, and $\lambda_3 = 3$. Here we would expect to choose $Q = 3$ from the data.
- $g_2(t) = \Psi_3(t)$.
- $\sigma = 1, \varepsilon_{ij}(t)$ and $\eta_{ij}(t)$ are taken to be Gaussian processes, because we are interested in hypotheses testing.

In this example, which we call Case 1, it can be shown that $Q \geq 3$ is needed to ensure that g_1 and g_2 are in the span of $\{\psi_1(t), \ldots, \psi_Q(t)\}$. For $Q = 3$, the optimal directions are $v_1^*(t) = g_1(t) = \Psi_1(t)$ and $v_2^*(t) = g_2(t) = \Psi_3(t)$. Thus an increase in power is not expected from using the optimal projection for testing. In an exploration of the power of the tests for H_0: $\tau_1 = 0$ and H_0: $\tau_2 = 0$, we set $a_1 = 1$ and $a_2 = .7$. Using (11), the tests of hypotheses concerning τ_1 and τ_2 should have approximately 50% and 55% power. In an investigation of the size of the tests, we set $a_1 = 0$ and $a_2 = 0$.

Next we considered an example where an increase in power could be expected from the use of the optimal projection. We call this example Case 2. The same design as just described was used except for the following: $\lambda_1 = 4, \lambda_2 = 3, \lambda_3 = 20$, and $g_2(t) = .9\Psi_2(t) + (1 - .9^2)^{1/2}\Psi_3(t)$. In this case it can be shown that $Q \geq 3$ is needed to ensure that g_1 and g_2 are in the span of $\{\psi_1(t), \ldots, \psi_Q(t)\}$. For $Q = 3, v_1^*(t) = g_1(t) = \Psi_1(t)$, but $v_2^*(t) = .9\Psi_2(t)(\lambda_2 + \sigma_2^2)^{-1} + (1 - .9^2)^{1/2}\Psi_3(t)(\lambda_3 + \sigma_3^2)^{-1} \neq g_2(t)$. Using (11), the tests of hypotheses concerning τ_1 and τ_2 based on $\tilde{W}_{ij}^{(1)}$ and $\tilde{W}_{ij}^{(2)}$ should have approximately 74% and 31% power. Using (12), the tests of hypotheses concerning τ_1 and τ_2 constructed from the optimal projections should have approximately 74% and 48% power.

Finally, we investigated the power and size of the tests for τ_1 and τ_2 based on the projections with the added complication of missing data. Each point (t_{ijk}, Y_{ijk}) was independently deleted with probability p_k. Here $(p_1, \ldots, p_7) = (0, 0, .03, .06, .09, .12, .15)$ was chosen to mimic the percentage missing in the Droloxifene study.

Table 1 lists the simulated size and power of the tests for a significance level of $\alpha = .05$. The results for the missing-data simulations are recorded within the parentheses. When $\tilde{Y}_{ij}(t)$ were projected along the directions $g_1(t), g_2(t)$, or $\hat{v}_1^*(t)$, a test of the hypothesis was deemed significant when the p value was less than $\alpha = .05$. When $\tilde{Y}_{ij}(t)$ were projected along the directions $\hat{g}_2(t)$ or $\hat{v}_2^*(t)$, a test of the hypothesis was deemed significant when the p value was less than $\alpha/(Q - 1) = .025$. The need for this Bonferroni-type adjustment was discussed in Section 4.

From these simulations, we see that the size of the test hovers about $\alpha = .05$ for testing $\tau_1 = 0$ in all cases. For testing $\tau_2 = 0$, the size of the test was inflated in small bandwidths ($b = .2$), especially when the data were projected onto $\hat{g}_2$. The size of the tests based on projecting on the estimated optimal projection were within an acceptable range. The size of the test is more sensitive to the bandwidth choice for the variable visit times. A bandwidth of .35 holds the size of the test near the significance level. The power for the tests of hypotheses concerning τ_1 and τ_2 is as predicted earlier using (11) and (12). Case 1 demonstrates that there is essentially no loss in power in using the estimated optimal projections when g_1 and g_2 are optimal in the fixed visit time case, and a slight loss of power in the variable time case. However, Case 2 demonstrates that there is an expected large increase in power in testing for $\tau_2 = 0$ with the optimal projection for both simulated visit time distributions. The sizes of the tests were found to increase with missing data, although increasing the bandwidth seems to alleviate the problem.

For comparison, a random-effects model (Laird and Ware 1982) was also fit to the simulated data using Proc Mixed in SAS. In the model for the mean $\mu(d, t)$, everything was assumed known except a_1 and a_2 which were estimated from the data. The hypotheses H_0: $\tau_i = 0$ are equivalent to H_0: $a_i = 0$ for $i = 1, 2$. The correct covariance structure with $V(s, t) = \sum_{i=1}^{3} \lambda_i \Psi_i(t)\Psi_i(s)$ was assumed, and only $\lambda_1, \ldots, \lambda_3$ and σ^2 were estimated from the data. Table 1, in the row labeled LW, lists the simulated size and power of the tests for a significance level of $\alpha = .05$ with complete and missing data. It can be seen that the sizes of the LW tests hover about $\alpha = .05$ in all cases. The power for testing hypotheses using the optimal projection given the nonparametric structured model compares favorably with the power for testing hypotheses using the LW random-effects model. This part of the simulations emphasizes the importance of the optimal projection for the structured model, although we note that the power for testing hypotheses using the optimal projection given the structured model can be very sensitive to bandwidth choice; see Case 1.

8. SIMULATIONS FOR THE GENERAL NONPARAMETRIC MODEL

Here we report the results of a small simulation study conducted to investigate the proposed method for the model given by (1) and (3). The same design used in the simulation study of Section 7 is used here. For $a_1 = 0$ and $a_2 = 0, A(s, t) = 0$, so all of the eigenvalues equal 0. The estimate $\hat{A}(s, t)$ will not necessarily have all its eigenvalues equal to 0. However, projecting $\tilde{Y}_{ij}(t)$ onto $\hat{g}_k(t)$ should reveal that $\tau_k = 0$ for $k \geq 1$.

For $a_1 = 1$ and $a_2 = .7$, there is only one nonzero eigenvalue for $A(s, t)$. For this nonparametric model, $g_1(t)$ is the eigenfunction corresponding to the eigenvalue (30/29)1.49:

$$g_1(t) = \begin{cases} (\psi_1(t) + .7\psi_3(t))/\sqrt{1.49} & \text{for Case 1,} \\ (\psi_1(t) + .63\psi_2(t) + .7(1 - .9^2)^{1/2}\psi_3(t))/\sqrt{1.49} & \text{for Case 2.} \end{cases}$$

Projecting $\tilde{Y}_{ij}(t)$ onto $\hat{g}_1(t)$ should reveal that $\tau_1 \neq 0$ with about 80% power in both cases according to (8) with $v = \hat{g}_1$ and $r = 1$. Projecting $\tilde{Y}_{ij}(t)$ onto $\hat{g}_k(t)$ should reveal that $\tau_k = 0$ for $k \geq 2$.

In the analysis of the simulated data, significance of the test statistic for testing H_0: $\tau_k = 0$ is declared if the p value $< \alpha/Q = \alpha/3$, because there are three directions of variation for the patient random effects. Table 1 reports the size and power for the test of hypothesis H_0: $\tau_1 = 0$ under the column heading $\hat{g}_1^+$ using the significance level $\alpha = .05$. In all of the simulated realizations, the hypothesis H_0: $\tau_2 = 0$ was not rejected. This is the desired result, because in fact $\tau_2 = 0$ in both cases. The results for the missing-data simulations are recorded within the parentheses.

Table 1. Size and Power of the One-Way ANOVA Based on the Projections; Q = 3

	Variable visit time						*Fixed visit time*					
	Testing τ_1 projection direction			*Testing τ_2 projection direction*			*Testing τ_1 projection direction*			*Testing τ_2 projection direction*		
	g_1	$\hat{v}_1^*$	$\hat{g}_1^+$	g_2	$\hat{g}_2$	$\hat{v}_2^*$	g_1	$\hat{v}*1$	$\hat{g}_1^+$	g_2	$\hat{g}_2$	$\hat{v}_2^*$
						Case 1						
$a_1 = 0, a_2 = 0$												
$b = .2$	.03 (.03)	.03 (.03)	.10 (.14)	.06 (.07)	.20 (.13)	.08 (.07)	.04 (.03)	.04 (.02)	.12 (.12)	.05 (.05)	.14 (.13)	.12 (.11)
$b = .35$	.04 (.03)	.04 (.04)	.07 (.07)	.04 (.03)	.03 (.06)	.03 (.04)	.04 (.03)	.04 (.02)	.09 (.07)	.04 (.02)	.05 (.07)	.05 (.07)
LW	.04 (.02)			.09 (.03)			.04 (.02)			.09 (.04)		
$a_1 = 1, a_2 = .7$												
$b = .2$	.53 (.63)	.53 (.63)	.83 (.85)	.61 (.58)	.63 (.64)	.55 (.58)	.52 (.63)	.53 (.64)	.85 (.85)	.63 (.62)	.65 (.66)	.62 (.61)
$b = .35$	.53 (.65)	.53 (.67)	.72 (.72)	.45 (.40)	.37 (.44)	.35 (.40)	.52 (.63)	.53 (.64)	.77 (.77)	.44 (.42)	.45 (.49)	.46 (.48)
LW	.51 (.61)			.63 (.63)			.52 (.60)			.62 (.63)		
						Case 2						
$a_1 = 0, a_2 = 0$												
$b = .2$	.05 (.04)	.04 (.04)	.10 (.11)	.08 (.09)	.11 (.07)	.12 (.08)	.03 (.04)	.03 (.04)	.08 (.05)	.07 (.09)	.08 (.06)	.09 (.10)
$b = .35$	.03 (.05)	.03 (.03)	.05 (.03)	.04 (.05)	.04 (.03)	.05 (.04)	.04 (.05)	.03 (.03)	.05 (.04)	.05 (.07)	.04 (.03)	.06 (.07)
LW	.04 (.02)			.03 (.05)			.04 (.02)			.04 (.05)		
$a_1 = 1, a_2 = .7$												
$b = .2$	.75 (.75)	.74 (.74)	.78 (.82)	.37 (.38)	.33 (.40)	.52 (.46)	.72 (.77)	.74 (.75)	.79 (.82)	.38 (.38)	.29 (.36)	.49 (.47)
$b = .35$	.75 (.76)	.74 (.74)	.74 (.79)	.36 (.34)	.30 (.27)	.46 (.44)	.74 (.76)	.68 (.73)	.77 (.79)	.35 (.36)	.26 (.31)	.44 (.37)
LW	.76 (.81)			.55 (.48)			.75 (.81)			.53 (.50)		

NOTE: Results for missing-data cases are given in parentheses g_i, $\hat{g}_i$, and $\hat{v}_i^*$ are the projections corresponding to the true g_i, estimated g_i, and optimal direction for $i = 1, 2$ in the nonparametric repeated-measures ANOVA model. $\hat{g}_1^+$ is the projection on the estimated g_1 in the general nonparametric model.

For comparison, the LW random-effects model fitted in the previous section is used. That model for the mean $\mu(d,t)$ cannot be rewritten as a linear model in terms of $g_1(t)$ defined earlier. The power for detecting deviations from the null hypotheses is greater for the general nonparametric model than for either the structured nonparametric repeated-measures ANOVA model or the random-effects LW model. The effect of the bandwidth and missing data on the tests using the nonparametric model are similar to that observed for the structured model. Specifically, the size of the test seems to increase with decreasing bandwidth for the variable visit times. For the most part, the level of the tests fell within an acceptable range. To hold the size of the test near the significance level when there are missing data, it is necessary to use a larger bandwidth.

9. DATA ANALYSIS

Here we analyze two datasets: The data from the Droloxifene Phase I clinical trial in patients with metastatic breast cancer (Buzdar et al. 1994) is in Section 9.1 and the data from the DFMO in vivo tumor growth experiment in Section 9.2. The bandwidths were selected by visual inspection of the plots. We did not smooth over the dose variable.

9.1 Droloxifene Clinical Trial

Figure 1 shows follicle-stimulating hormone (FSH) levels for each woman in the clinical trial plotted by dose group. Approximately 7%, 3%, 3%, and 14% of the FSH measurements are missing at each of the four targeted time points. The final frame shows $\hat{\mu}(d,t)$ for the five dose groups. These smooths use a bandwidth of $b = .4$ on the standardized time scale. A value of Q must be selected to begin the analysis. The first five eigenvalues of $\hat{\mathbf{V}}(s,t)$ are 643, 15, .7, .06, and .002. The first two eigenvalues explain more than 99% of the total variation of $\hat{\mathbf{V}}(s,t)$. Thus $Q = 2$ was selected to construct the smooths $\tilde{Y}_{ij}(t)$ shown in Figure 1. Plots of the centered principal component scores $\hat{a}_{ijq} - \hat{a}_{i\cdot q}$ for $q = 1, 2$ versus age or baseline FSH did not reveal any dependence of the patient random effects on those covariates.

We analyze the data using the nonparametric model given by (1) and (3). The first three eigenvalues of $\hat{A}(s,t)$ are 109, 29, and 4, and the corresponding cumulative eigenvalues account for 77%, 97%, and 99.97% of the total variation in the dose group means. The first three eigenfunctions of $\hat{A}(s,t)$ correspond to the estimates of $\hat{g}_1(t), \hat{g}_2(t)$, and $\hat{g}_3(t)$ in the nonparametric model for $\mu(d,t)$. For these data, the shapes of these three functions correspond to roughly constant, linear, and quadratic functions. The p values for testing H_0: $\tau_k = 0$ for $k = 1, \ldots, 7$ are .3, .004, .001, 1, 1, 1, 1. Thus τ_2 and τ_3 are significantly different from 0(p value $< \alpha/Q = .05/2$), and we would not expect the structured model to hold for these data. Figure 2 plots the projections of $\tilde{Y}_{ij}(t)$ in the direction of $\hat{g}_1(t), \hat{g}_2(t)$, and $\hat{g}_3(t)$. Treating dose as a continuous variable, a kernel smooth of the projections in each scatterplot estimates $\hat{f}_1(d), \hat{f}_2(d)$, and $\hat{f}_3(d)$. The kernel estimates use a bandwidth of .4 on the standardized dose scale. Note that the τ's are not estimated from the $\hat{f}$'s, but rather are estimated using (13). Inspection of Figure 2 reveals that the significance of τ_2 and τ_3 is likely not attributable to a dose–response effect, because no consistent trend emerged in the scatterplots for $\hat{f}_2(d)$ and $\hat{f}_3(d)$. However, closer inspection of the second frame in Figure 2 reveals that all six patients in the highest dose

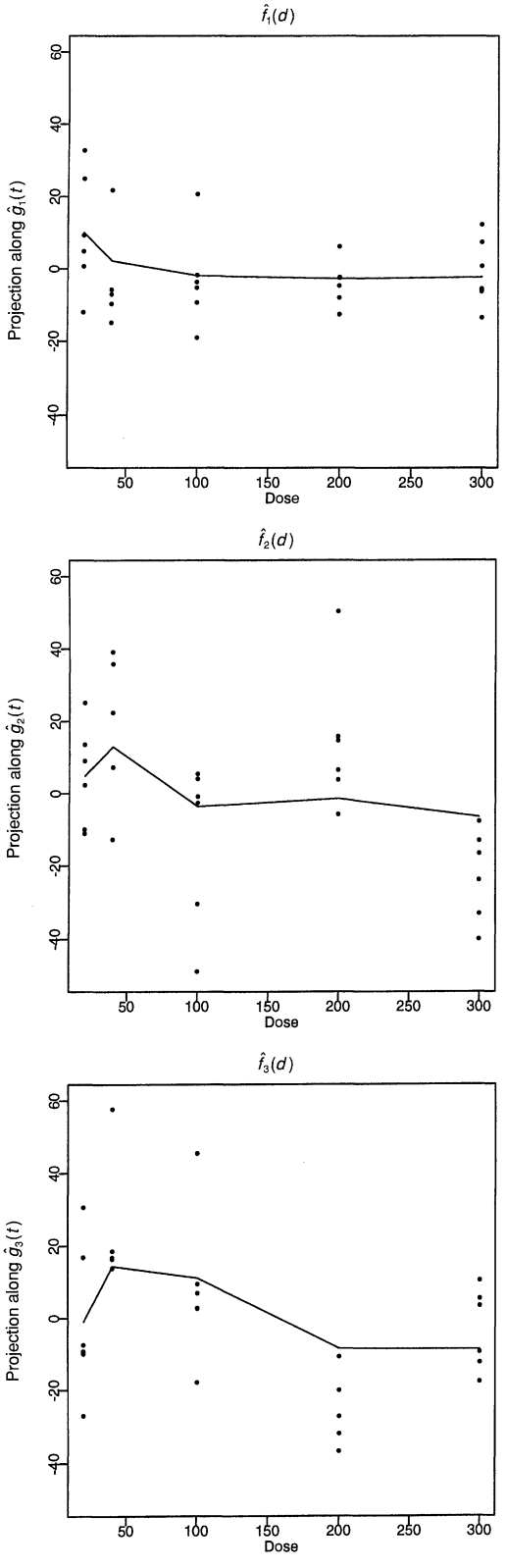

Figure 2. General Nonparametric Model for the Droloxifene Data: $\hat{f}_1$, $\hat{f}_2$, and $\hat{f}_3$ Versus Dose. There is no support for a dose–response effect because the estimated function values appear to randomly vary about 0.

group had a negative slope over time, as described by $\hat{f}_2(d)\hat{g}_2(t)$. This suggests that the highest dose of Droloxifene depresses FSH levels over time.

The parameter τ_1 was not found to be significantly different from 0 using a one-way ANOVA, although, a smooth pattern as a function of dose emerges from the first frame in Figure 2. This pattern is most likely due to significantly higher baseline FSH levels for women in the first dose group.

A plot of $\hat{h}(t)$ (not shown) suggests that on the average, the women in the study experienced a drop of about 10 units of FSH level over the first 2–3 weeks of the study, followed by a stabilization of FSH level. Because there was no control group, it cannot be determined from these data; whether this drop should be attributed to Droloxifene or to disease progression.

In summary, there is no evidence of a dose–response effect on FSH level using the nonparametric model. There was evidence that the patients in the highest dose group did experience a 20-unit drop in FSH level over time, whereas the other patients experienced a 10-unit drop in FSH level over time. The same conclusions were reached independently of bandwidth choice.

9.2 DFMO In Vivo

Figure 3 shows a plot of the tumor log-volume versus time for each mouse by dose group. Approximately 3%, 3%, 3%, 7%, 8%, and 11% of tumor volumes are missing at each of the six targeted time points. The final frame shows $\hat{\mu}(d,t)$ for the five dose groups. These smooths use a bandwidth of $b = .25$ on the standardized time scale. The first three eigenvalues of $\hat{V}(s,t)$ are .5, .01, and .008. The first two eigenvalues explain more than 98% of the total variation of $\hat{\mathbf{V}}(s,t)$. The third eigenfunction looked more like noise than a smooth direction for an animal's tumor log-volume profile random effect. Hence $Q = 2$ was selected.

First, we analyzed the data using the nonparametric model given by (1) and (3). The first three eigenvalues of $\hat{A}(s,t)$ are .1, .001, and .0001. The first two eigenvalues account for more than 99% of the total variation in the dose group means. The p values for testing H_0: $\tau_k = 0$ for $k = 1, \ldots, 7$ are .003, .9, 1, 1, 1, 1, 1. Only τ_1 is significantly different from 0 (p value $< \alpha/Q = .05/2$). It is appropriate to try to fit the structured model given by (1) and (2).

Figure 4 shows an estimate of $g_2(t)$ for the structured model for the mean given by (1) and (2). This estimate of $g_2(t)$ from the structured model agrees with the estimate of $\hat{g}_1(t)$ (not shown) obtained from fitting the general nonparametric model. This curve suggests that the tumor log-volume may start out linear in time but later flattens out. Assumption A does not appear to be completely satisfied, although a large component of g_1 and g_2 lies in the span of $\hat{\Psi}_1$ and $\hat{\Psi}_2$. The appropriateness of using the optimal projections as provided by Theorem 1 under assumption A is suspect.

The second frame of Figure 4 shows a plot of the projection of the smoothed data $\tilde{Y}_{ij}(t)$ in the direction $\hat{g}_2(t)$. Treating the dose as a continuous variable, superimposed on that projection is a kernel estimate $\hat{f}_2(d)$ obtained by smoothing that scatterplot using the bandwidth .40 on the standardized dose scale. The shape of $\hat{f}_2(d)$ is not linear, and it is doubtful that a quadratic function in dose would capture the shape of $\hat{f}_2(d)$, because it levels off so quickly after a sharp de-

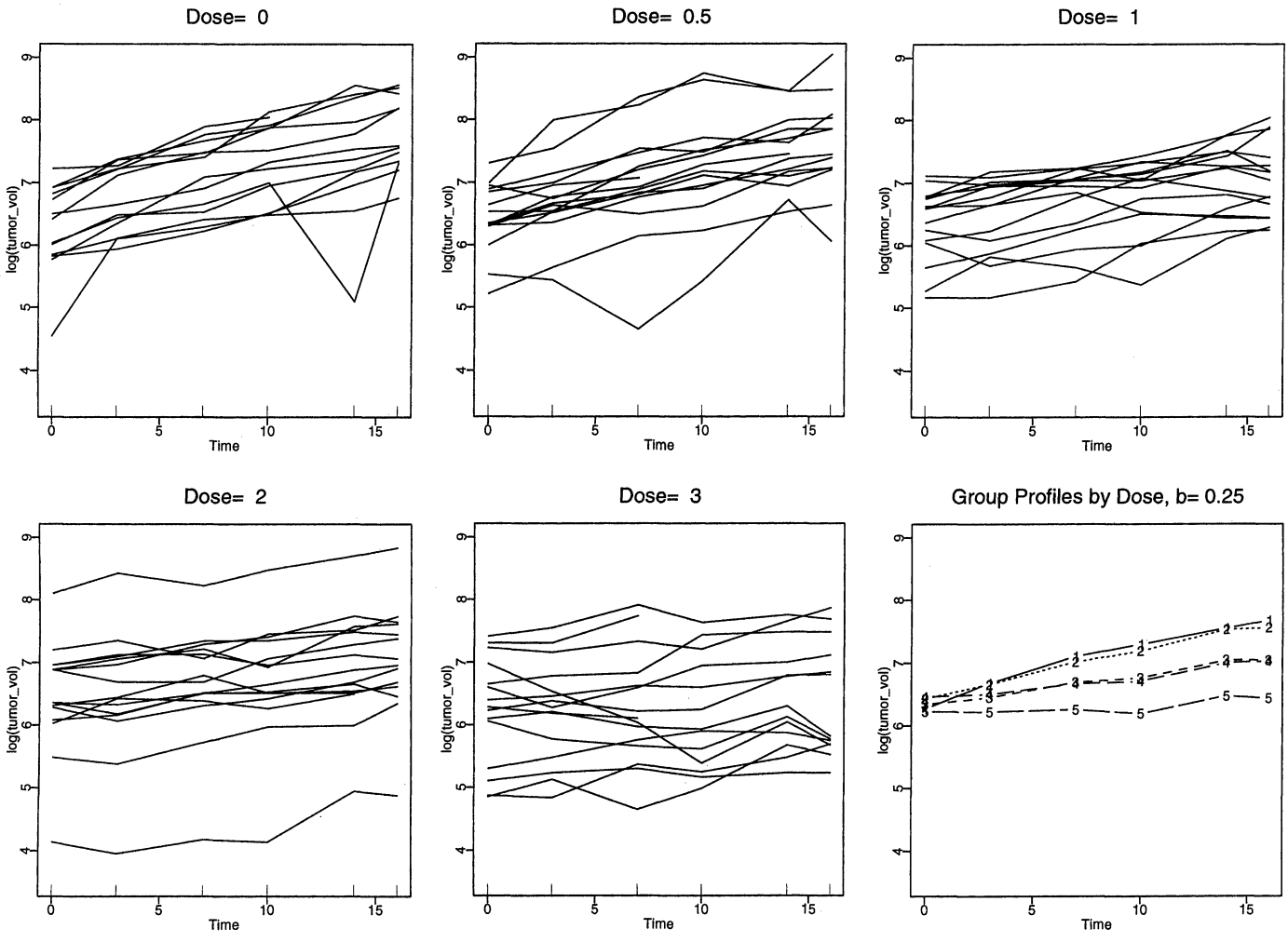

Figure 3. Animal Profiles of DFMO In Vivo Experiment (Raw Data Connected by Solid Lines). Group profiles by dose are obtained by smoothing data aggregated by group. 1 ——, dose = 0; 2 · · ·, dose = .5; 3 - - -, dose = 1; 4 – – –, dose = 2; 5 — — —, dose = 3.

cline. A test of H_0: $\tau_2 = 0$ was conducted using a one-way ANOVA (p value $= 1.87 \times 10^{-12} < \alpha/(Q-1) = .05$) of the projections $\tilde{W}_{ij}^{(2)}$; it was significant. Recall that $\hat{\tau}_2$ is obtained from (5) and not $\hat{f}_2(d)$.

A plot of $\hat{h}(t)$ suggests that on average, the animals experienced a linear growth over time in tumor volume from 6.4 to 7.2 units. The structured model suggests that the mean tumor volume at $t = 16$ is expected to be decreasing according to 7.7, 7.6, 7.0, 6.9, and 6.7 as the DFMO dose increases from 0 to 3. It is interesting to note that although assumption A was suspect for this dataset, the same conclusions are reached using the optimal projections. In summary, the data suggest that DFMO has a dose effect on the tumor log-volume. Higher doses of DFMO are associated with a smaller rate of tumor growth over time. These conclusions were supported by both the nonparametric and structured model, which for this dataset were in complete agreement. There was no evidence for lack of fit of the structured model.

Next, we fitted a parametric random effects model to the data using Proc Mixed in SAS. The eigenfunctions of $\hat{\mathbf{V}}(s,t)$ suggested that the random effects be included in the model as multiples of 1, t and t^2. The 3×3 matrix ϕ for the covariance of the random effects was left unstructured and was estimated from the data. The covariance matrix $\mathbf{R}$ for the error term was taken to be an unknown multiple of the identity. First (as in Heitjan et al. 1993), the model for the fixed effects was fitted with a t, d, and $d \times t$ with dose included as a categorical variable. This model allows for each dose group to have an arbitrary slope describing the dependence of the response on time. It is similar to the nonparametric structured model in that we did not smooth over the dose variable. Dose by itself was not significant, but the dose-by-time interaction was significant. A plot of the arbitrary slopes versus dose agreed with the shape of $\hat{f}_2$ in Figure 4. Heitjan et al. (1993) tried to obtain a simpler parametric description of the dependence of the arbitrary slopes on dose. They considered several parametric models with dose as a continuous variable, including a linear model and a quadratic model in dose for the interaction. Heitjan et al. (1993, p. 6045) concluded that "none fit as well as the model with arbitrary doses." Nevertheless, we fitted another parametric random-effects model with dose as a continuous variable. The fixed effect was chosen to have an intercept, $t, d, d^2, d \times t$, and $d^2 \times t$. Only the intercept, the t and $d \times t$ terms, were highly significant (p value $< .0001$). The parameter estimate for the $d^2 \times t$ by time

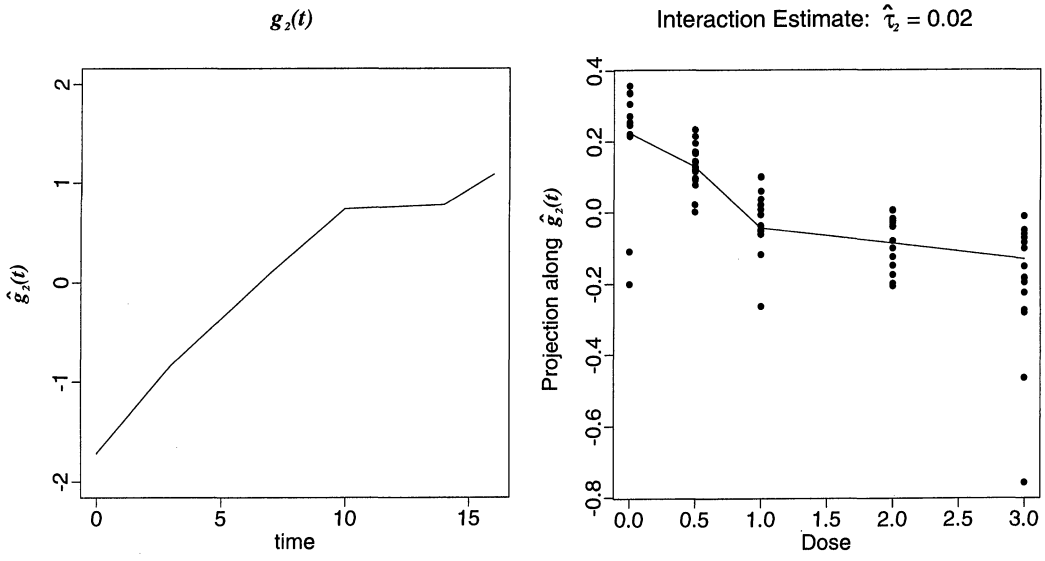

Figure 4. Structured Nonparametric Model for the DFMO Data: $\hat{g}_2(t)$ and $\hat{f}_2(d)$. The function $\hat{g}_2(t)$ suggests that the tumor log-volume may start out linear in time, but later flattens out. The function $\hat{f}_2(d)$ suggests that the tumor log-volume rate of growth depends smoothly on dose. At low doses, the rate of growth declines sharply with increasing dose. Decline in the rate of growth levels off at the higher doses.

interaction was positive and marginally significant (p value of .034). This supports our earlier findings with the structured model that $f_1(d)$ is not significantly different from 0 and that $\hat{f}_2(d)$ has some curvature. A reduced parametric model with an intercept, t, d, and $d \times t$ for fixed effects, was also fitted to the data. The restricted maximum likelihood (REML) for the reduced model is greater than the REML of the larger model. This may be due to the marginal significance of the $d^2 \times t$ interaction and the penalty of estimating two additional parameters. A satisfactory parametric model for the fixed effects is not forthcoming and was not pursued any further.

10. CONCLUSIONS

We have presented methods for exploratory data analysis as a step in building a parametric model for repeated-measures data. Guidance for selection of both the covariance structure and the mean effect were provided. In the case where a parametric model is either not forthcoming or needed, methods for inference about the mean effects were developed using both a structured nonparametric repeated-measures ANOVA model and a general nonparametric model. Bandwidths were chosen by visual inspection. [See Staniswalis and Lee (1997) for suggestions for data based bandwidth selection procedures.] Further work is needed on criteria for the selection of bandwidth and Q from the data.

Through simulations, we demonstrated the usefulness of the approximations for nonparametrically testing hypotheses addressing the presence of a dose–shift effect or dose-by-time interaction. Missing data and variation in the sampling points over time among patients were easily handled by this approach using nonparametric smoothing. Although dose was a discrete variable for both of the datasets analyzed and for the simulated data generated in the Monte Carlo study, we developed the methodology for the more general situation when the variable "dose" is a vector of explanatory variables. An S-PLUS implementation of this methodology which was used for the simulations and data analysis is available from the second author or STATLIB (http://lib.stat.cmu.edu).

APPENDIX: PROOFS

Choice of Quadrature Weights

Because $\{\mathbf{\Psi}_{ij1}, \ldots, \mathbf{\Psi}_{ijQ}\}$ are not necessarily orthonormal, a diagonal matrix $\mathbf{W}_{ij}$ is needed that satisfies the $Q(Q+1)/2$ constraints given by $\mathbf{\Psi}_{ijs}^t \mathbf{W}_{ij} \mathbf{\Psi}_{ijq} = 1$ if $q = s$ and 0 otherwise, for $q \leq s = 1, \ldots, Q$. Let $\mathbf{w}_{ij}$ denote the vector of diagonal elements of $\mathbf{W}_{ij}$. Then $\mathbf{w}_{ij}$ can be obtained as a solution to a linear system of equations of the form $\mathbf{A}_{ij}\mathbf{w}_{ij} = \mathbf{b}$, where the $Q(Q+1)/2$ rows of $\mathbf{A}_{ij}$ are given by the diagonal of $\mathbf{\Psi}_{ijq}\mathbf{\Psi}_{ijs}^t$. The vector $\mathbf{b}$ is just a column of 0s and 1s corresponding to the constraints. If it exists, then the solution $\mathbf{w}_{ij}$ can be obtained using the singular value decomposition of $\mathbf{A}_{ij}$ (Horn and Johnson 1990, p. 429, eq. 7.4.5). If a solution does not exist, then the least squares solution is similarly obtained.

A.1 Proof of Theorem 1

At most, it can be said that if $v(t) \in L_2(P)$, then $v(t) = v_0(t) + \lim_{N\to\infty} \sum_{i=1}^N \langle v, \Psi_i \rangle \Psi_i(t)$, where the "lim" means the limit in the $L_2(P)$ sense and v_0 satisfies $\int_0^1 V(s,t) v_0(s)\, dP(s) = 0$ (Griffel 1981, prop. 9.3, p. 242). Thus under assumption (A), we need only minimize (9) subject to (6) among all $v \in L_2(P)$ satisfying

$$v(t) = \lim_{N\to\infty} \sum_{i=1}^{N} v_i \Psi_i(t), \tag{A.1}$$

where $v_i = \langle v, \Psi_i \rangle$.

The variance (9) of the data given in the scatterplot is the quantity to be minimized subject to (6) and (14). Note that (9) is equivalent to

$$\frac{\sum_{i=1}^{Q}(\lambda_i+\sigma_i^2)v_i^2}{\sum_{i=1}^{Q} v_i a_{ri}}, \qquad r=1,2.$$

Thus, under assumption A, we can assume without loss of generality that $v_i^* = 0$ for $i > Q$. All of the summations in the development that follows run over $1,\ldots,Q$. Set $\bar{\lambda}_i = \lambda_i + \sigma_i^2$.

We approach the problem by showing that $v_r^*(t)$ is a minimizer of

$$H(\mathbf{v},\lambda) = \frac{\sum_i \bar{\lambda}_i v_i^2}{(\sum_i v_i a_{ri})^2} + \lambda \sum_i v_i a_{r'i}$$

using the Lagrange multiplier method. This will establish the result, because the minimizer $v_r^*(t)$ in fact satisfies the constraints (6) and (A.1).

Proceed by computing $[\partial H(\mathbf{v},\lambda)]/\partial v_k$ and setting it equal to 0. Using the fact that $\int_0^1 g_{r'}^2(t)\,dP(t) = \sum_k a_{r'k}^2 = 1$, and $\int_0^1\int_0^1 g_r(t)g_{r'}(t)\,dP(s)\,dP(t) = \sum_k a_{rk}a_{r'k} = 0$, we obtain

$$\lambda = \frac{-2(\sum_k a_{r'k}\bar{\lambda}_k v_k)}{(\sum_i v_i a_{ri})^2}.$$

Now plug in the solution for λ into $[\partial H(\mathbf{v},\lambda)]/\partial v_k = 0$, get a common denominator and simplify. Verify that v_{rk}^* satisfies the equation obtained.

Note that any multiple of $v_r^*(t)$ given by the theorem will also have generalized Fourier coefficients that satisfy this equation. A unique solution is given by placing an additional constraint, such as $\sum_i v_i a_{ri} = 1$, on $v_r^*(t)$. In this case we are minimizing $\sum_i \bar{\lambda}_i v_i^2$ with $\bar{\lambda}_i > 0$ subject to two linear constraints, $\sum_i v_i a_{ri} = 1$ and $\sum_i v_i a_{r'i} = 0$. The unique minimizer is the point closest to the origin in the norm $\sum_i \bar{\lambda}_i v_i^2$ that lies in the intersection of the hyperplane defined by the two linear constraints.

Proof of Corollary 1

The covariance $V(s,t)$ of the stochastic process $U_1 g_1(t) + U_2 g_2(t)$ is given by

$$[g_1(t) \quad g_2(t)]\mathbf{\Phi}\begin{bmatrix} g_1(s) \\ g_2(s)\end{bmatrix} = [g_1(t) \qquad g_2(t)]\mathbf{P}\mathbf{\Lambda}\mathbf{P}^t\begin{bmatrix} g_1(s) \\ g_2(s)\end{bmatrix},$$

where $\mathbf{P}$ is an orthogonal 2×2 matrix and $\mathbf{\Lambda}$ is 2×2 diagonal matrix of the eigenvalues of $\mathbf{\Phi}$ such that $\mathbf{\Phi} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^t$. Because $\mathbf{P}$ is orthogonal, it must be of the form

$$\mathbf{P} = \begin{bmatrix} a & -b \\ b & a\end{bmatrix},$$

where $a^2 + b^2 = 1$. The eigenfunctions of $V(s,t)$ and $C(s,t)$ are given by $\phi_1(t) = a g_1(t) + b g_2(t)$ and $\phi_2(t) = -b g_1(t) + a g_2(t)$. Thus $g_1(t) = a\phi_1(t) - b\phi_2(t)$ and $g_2(t) = b\phi_1(t) + a\phi_2(t)$. So we have $a_{11} = a, a_{12} = -b, a_{21} = b$, and $a_{22} = a$. From an application of the theorem, we can then show that $v_{ri}^* = a_{ri}$.

Proof of Theorem 2

The proof is written out for the case where $n_i \to \infty, i = 1,\ldots,m$ and $s \neq t$. The proof is easily modified when this is not satisfied and is not included here. The assumptions that are sufficient to show that both $\hat{\mu}(s)$ and $\hat{\mu}(d,s)$ are uniformly consistent estimators of $\mu(s)$ and $\mu(d,s)$ also yield the desired result once we are able to establish that $E[\hat{C}(s,t)] \to C(s,t)$ as $n\to\infty$. Here the expected value is taken conditionally on the visit times $\{t_{ijk}\}$. Now $E[\hat{C}(s,t)]$ depends on $E[(Y_{ijk} - \hat{\mu}(d_i,t_{ijk}))(Y_{ijk'} - \hat{\mu}(d_i,t_{ijk'}))] = E[(Y_{ijk} - \hat{\mu}_i(t_{ijk}))(Y_{ijk'} - \hat{\mu}_i(t_{ijk'}))] + o(1)$ by using the Cauchy–Schwarz inequality.

Consider the four terms that arise in $E[(Y_{ijk} - \hat{\mu}_i(t_{ijk}))(Y_{ijk'} - \hat{\mu}_i(t_{ijk'}))]$;

$$A_1 = E[Y_{ijk}Y_{ijk'}] = C(t_{ijk},t_{ijk'}) + \mu_i(t_{ijk})\mu_i(t_{ijk'}),$$

$$\begin{aligned} A_2 &= E[Y_{ijk}\hat{\mu}_i(t_{ijk'})] \\ &= \sum_r C(t_{ijk},t_{ijr})\frac{w(t_{ijk'},t_{ijr};b)}{\sum_q\sum_r w(t_{ijk'},t_{iqr};b)} \\ &\quad + \mu_i(t_{ijk})E[\hat{\mu}_i(t_{ijk'})], \\ A_3 &= E[Y_{ijk'}\hat{\mu}_i(t_{ijk})] \\ &= \sum_r C(t_{ijk'},t_{ijr})\frac{w(t_{ijk},t_{ijr};b)}{\sum_q\sum_r w(t_{ijk},t_{iqr};b)} \\ &\quad + \mu_i(t_{ijk'})E[\hat{\mu}_i(t_{ijk})], \end{aligned}$$

and

$$\begin{aligned} A_4 &= E[\hat{\mu}_i(t_{ijk})\hat{\mu}_i(t_{ijk'})] \\ &= \frac{\sum_q\sum_r\sum_h w(t_{ijk},t_{iqr};b)w(t_{ijk'},t_{iqh};b)C(t_{iqr},t_{iqh})}{[\sum_q\sum_r w(t_{ijk},t_{iqr};b)][\sum_q\sum_h w(t_{ijk'},t_{igh};b)]} \\ &\quad + E[\hat{\mu}_i(t_{ijk})]E[\hat{\mu}_i(t_{ijk'})]. \end{aligned}$$

Then, using standard results from smoothing, it can be shown that

$$A_1 - A_2 - A_3 + A_4 = C(t_{ijk},t_{ijk'})c_{ij}(t_{ijk},t_{ijk'}) + O(b) + O(\min\{n_i^{-1},K^{-1}\}),$$

where K is the least upper bound on the number of clinic visits per patient. Note that $c_{ij}(t_{ijk},t_{ijk'})$ is defined in Section 6 and is $O(1)$. Thus $(A_1 - A_2 - A_3 + A_4)c_{ij}^{-1}(t_{ijk},t_{ijk'})$ is approximately equal to $C(t_{ijk},t_{ijk'})$ and $E[\hat{C}(s,t)] = C(s,t) + o(1)$.

[Received February 1997. Revised March 1998.]

REFERENCES

Altman, N. S. (1990), "Kernel Smoothing of Data With Correlated Errors," *Journal of the American Statistical Association*, 85, 749–759.

Altman, N. S., and Casella, G. (1995), "Nonparametric Empirical Bayes Growth Curve Analysis," *Journal of the American Statistical Association*, 90, 508–515.

Berkey, C. S., and Kent, R. L., Jr. (1983), "Longitudinal Principal Components and Non-Linear Regression Models of Early Childhood Growth," *Annals of Human Biology*, 10, 523–536.

Berkey, C. S., Laird, N. M., Valadian, I., and Gardner, J. (1991), "Modelling Adolescent Blood Pressure Patterns and Their Prediction of Adult Pressures," *Biometrics*, 47, 1005–1018.

Besse, P., and Ramsay, J. O. (1986), "Principal Components Analysis of Sampled Functions," *Psychometrika*, 51, 285–311.

Buckley, M. J. (1991), "Detecting a Smooth Signal: Optimality of Cusum-Based Procedures," *Biometrika*, 78, 253–262.

Buzdar, A. U., Kau, S., Hortobagyi, G. N., Theriault, R. L., Booser, D., Holmes, F. A., Walters, R., and Krakoff, I. H. (1994), "Phase I Trial of Droloxifene in Patients With Metastatic Breast Cancer," *Cancer Chemotherapy and Pharmacology*, 33, 313–316.

Castro, P. E., Lawton, W. H., and Sylvestre, E. A. (1986), "Principal Modes of Variation for Processes With Continuous Sample Curves," *Technometrics*, 28, 329–337.

Cheng, K. F., and Lin, P. E. (1981), "Nonparametric Estimation of a Regression Function," *Seitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57, 223–233.

Cox, D., Koh, E., Wahba, G., and Yandell, B. S. (1988), "Testing the (Parametric) Null Model Hypothesis in (Semiparametric) Partial and Generalized Spline Models," *The Annals of Statistics*, 16, 113–119.

Eubank, R. L. (1988), *Smoothing Splines and Nonparametric Regression*, New York: Marcel Dekker.

Eubank, R. L., and Hart, J. D. (1993), "Commonality of Cusum, von Neumann and Smoothing-Based Goodness-of-Fit Tests," *Biometrika*, 80, 89–98.

Eubank, R. L., Hart, J. D., Simpson, D. G., and Stefanski, L. A. (1995), "Testing for Additivity in Nonparametric Regression," *The Annals of Statistics*, 23, 1896–1920.

Fraiman, R., and Meloche, J. (1994), "Smoothing Dependent Observations," *Statistics and Probability Letters*, 21, 203–214.

Griffel, D. H. (1981), *Applied Functional Analysis*, New York: Ellis Horwood.

Hart, J. D., and Wehrly, T. E. (1986), "Kernel Regression Estimation Using Repeated Measurements Data," *Journal of the American Statistical Association*, 81, 1080–1088.

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, London: Chapman and Hall.

Heitjan, D. F., Manni, A., and Santen, R. J. (1993), "Statistical Analysis of *in Vivo* Tumor Growth Experiments," *Cancer Research*, 53, 6042–6050.

Heuser, H. G. (1982), *Functional Analysis*, New York: Wiley.

Horn, R. A., and Johnson, C. R. (1990), *Matrix Analysis*, New York: Cambridge University Press.

Jones, M. C., Davies, S. J., and Park, B. U. (1994), "Versions of Kernel-Type Regression Estimators," *Journal of The American Statistical Association*, 89, 825–832.

Jones, M. C., and Rice, J. A. (1992), "Displaying the Important Features of Large Collections of Similar Curves," *The American Statistician*, 46, 140–145.

Laird, N. M., Donnelly, C., and Ware, J. H. (1992), "Longitudinal Studies With Continuous Responses," *Statistical Methods in Medical Research*, 1, 225–247.

Laird, N. M., and Ware, J. A. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

Müller, H. G. (1988), *Nonparametric Regression Analysis of Longitudinal Data* (Lecture Notes in Statistics Vol. 46), New York: Springer-Verlag.

Nadaraya, E. A. (1989), *Nonparametric Estimation of Probability Densities and Regression Curves*, Dordrecht, The Netherlands: Kluwer.

Ramsay, J. O., and Dalzell, C. J. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society*, Ser. B, 53, 539–572.

Rice, J. A. (1984), "Boundary Modification for Kernel Regression," *Communications in Statistics, Part A—Theory and Methods*, 13, 893–900.

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*, Ser. B, 53, 233–243.

Robinson, G. K. (1991), "That BLUP Is a Good Thing: The Estimation of Random Effects," *Statistical Science*, 6, 15–51.

Silverman, B. W. (1996), "Smoothed Functional Principal Components by Choice of Norm," *The Annals of Statistics*, 24, 1–24.

Staniswalis, J. G., and Lee, J. J. (1997), "Nonparametric Regression Analysis of Longitudinal Data," Technical Report Series of the Statistical Consulting Laboratory, No. 28. University of Texas at El Paso, Dept. of Mathematical Sciences.

Staniswalis, J. G., and Messer, K. (1996), Addendum to "Kernel Estimators for Multiple Regression," *Journal of Nonparametric Statistics*, 7, 67–68.

Staniswalis, J. G., Messer, K., and Finston, D. R. (1994), "Kernel Estimators for Multiple Regression," *Journal of Nonparametric Statistics*, 3, 103–121.

Zeger, S. L., and Diggle, P. J. (1994), "Semiparametric Models for Longitudinal Data With Application to CD4 Cell Numbers in HIV Seroconverters," *Biometrics*, 50, 689–699.