Diagnostics for Linear Models With Functional Responses

Qing SHEN

Edmunds.com Inc.
Santa Monica, CA 90404
(*shenqing26@hotmail.com*)

Hongquan XU

Department of Statistics
University of California
Los Angeles, CA 90095
(*hqxu@stat.ucla.edu*)

Linear models where the response is a function and the predictors are vectors are useful in analyzing data from designed experiments and other situations with functional observations. Residual analysis and diagnostics are considered for such models. Studentized residuals are defined, and their properties are studied. Chi-squared quantile–quantile plots are proposed to check the assumption of Gaussian error process and outliers. Jackknife residuals and an associated test are proposed to detect outliers. Cook's distance is defined to detect influential cases. The methodology is illustrated by an example from a robust design study.

KEY WORDS: Cook's distance; Functional F test; Functional regression analysis; Jackknife residual; Studentized residual.

1. INTRODUCTION

Functional data analysis becomes increasingly popular as modern technology allows relatively easy access to data collected continuously over a period of time. Because only a finite number of observations are recorded, traditional multivariate analysis or longitudinal data analysis strategies may be applied. However, it is often more natural and convenient to think of, model, and analyze the observed functional data as single elements, rather than merely a sequence of individual (repeated) observations, as pointed out by Ramsay and Dalzell (1991). Ramsay and Silverman (2002, 2005) provided introductions to various issues on functional data analysis and applications in criminology, meteorology, medicine, and many other fields.

Linear models are useful in describing and predicting responses by a set of predictors. In a functional linear model, the response, the predictors, or both could be functions. In this article we assume that $\mathbf{Y}(t) = \mathbf{X}\boldsymbol{\beta}(t) + \boldsymbol{\epsilon}(t)$, where the responses are functions and the predictors are scalar vectors. Such linear models, including functional analysis of variance as a special case, have been studied by many authors (see, e.g., Ramsay and Silverman 2005; Faraway 1997; Fan and Lin 1998; Eubank 2000; Shen and Faraway 2004). Ramsay and Silverman (2005) laid out some general ideas on estimation and provided preliminary methods for inference. Faraway (1997) pointed out the inappropriateness of traditional multivariate test statistics, proposed a bootstrap-based testing method, and discussed residual analysis for such models. Fan and Lin (1998) proposed adaptive transform-based tests for functional analysis of variance models. Eubank (2000) considered tests for a constant mean function using a cosine basis function approach. Shen and Faraway (2004) recently proposed a functional F test for comparing nested functional linear models. We point out that much data from designed experiments fit well with such functional linear models; for example, Nair, Taam, and Ye (2002) used functional linear models to analyze some data from robust design studies.

Outliers and influential cases are frequently found not only in observational studies, but also in designed experiments. Including them in the analysis often leads to misleading conclusions. However, diagnostics have been largely ignored for functional linear models in the literature, most possibly due to lack of proper tools. Because no formal procedures are available, the current practice in functional regression analysis is to skip diagnostics completely or to detect outliers visually; for example, Faraway (1997) and Shen and Faraway (2004) removed an obvious outlier from their data. The purpose of this article is to formally define diagnostic statistics for functional linear regression models with fixed covariates and provide simple computational methods making it possible to automate the diagnostic checking procedure.

In Section 2 we briefly review functional linear models and a functional F test proposed by Shen and Faraway (2004). In Section 3 we define studentized residuals for residual analysis and propose chi-squared quantile–quantile (Q–Q) plots to check the assumption of Gaussian error process and outliers. We then define jackknife residuals and develop a formal test to detect outliers. We also define Cook's distance to detect influential cases. Formulas for easy computation of jackknife residuals and Cook's distance are given. We illustrate the methodology with an example from a robust design study in Section 4. We give concluding remarks in Section 5.

2. FUNCTIONAL LINEAR MODELS

Suppose that we have functional response data $y_i(t), i = 1, \ldots, n, t \in [a, b]$. We are interested in building a regression model for relating this response to a vector of predictors, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T$. The model takes the familiar form

$$y_i(t) = \mathbf{x}_i^T \boldsymbol{\beta}(t) + \epsilon_i(t), \tag{1}$$

where $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_p(t))^T$ are unknown parameter functions and $\epsilon_i(t)$ is a Gaussian stochastic process with mean 0 and covariance function $\gamma(s, t)$. We assume that $\epsilon_i(\cdot)$ and $\epsilon_j(\cdot)$ are independent for $i \neq j$.

The unknown functions $\boldsymbol{\beta}(t)$ can be estimated by minimizing $\sum_{i=1}^n \|y_i - \mathbf{x}_i^T \boldsymbol{\beta}\|^2$, where $\|f\| = (\int f(t)^2\, dt)^{1/2}$ is the L_2

TECHNOMETRICS, FEBRUARY 2007, VOL. 49, NO. 1
DOI 10.1198/004017006000000444

norm of function $f = f(t)$. This minimization leads to the least squares estimates $\hat{\boldsymbol{\beta}}(t) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}(t)$, where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is the usual $n \times p$ model matrix, whereas $\mathbf{Y}(t) = (y_1(t), \ldots, y_n(t))^T$ is a vector of responses. The predicted (or fitted) responses are $\hat{y}_i(t) = \mathbf{x}_i^T\hat{\boldsymbol{\beta}}(t)$, and the residuals are $\hat{\epsilon}_i(t) = y_i(t) - \hat{y}_i(t)$. The residual sum of squares is $rss = \sum_{i=1}^n \|\hat{\epsilon}_i\|^2 = \sum_{i=1}^n \int \hat{\epsilon}_i(t)^2\,dt$.

An important inference problem is to compare two nested linear models, ω and Ω, where ω has q parameter functions, Ω has p parameter functions, and ω results from a linear restriction on the parameter functions of Ω. A naive approach is to examine the pointwise t statistics on each time point for testing $\boldsymbol{\beta}(t)$. This entails a serious problem with multiple comparison; if Bonferroni corrections were applied to the significance level, then power would be significantly compromised, because the responses are often highly correlated within each unit. Faraway (1997) and Ramsey and Silverman (2005) proposed bootstrap-based and permutation-based tests, which require intensive computation. As pointed out by Faraway (1997), traditional multivariate test statistics are inappropriate because of the influence of unimportant variation directions.

To overcome these issues, Shen and Faraway (2004) proposed a functional F test. Define

$$F = \frac{(rss_\omega - rss_\Omega)/(p-q)}{rss_\Omega/(n-p)}, \tag{2}$$

where rss_ω and rss_Ω are residual sum of squares under models ω and Ω. When the null model is true, this functional F statistic is distributed like a ratio of two linear combinations of infinite independent chi-squared random variables, that is, $[(p-q)^{-1}\sum_{k=1}^\infty \lambda_k\chi_k^2(p-q)]/[(n-p)^{-1}\sum_{k=1}^\infty \lambda_k\chi_k^2(n-p)]$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are eigenvalues of the covariance function $\gamma(s,t)$, $\chi_k^2(a)$ is a chi-squared random variable with a degrees of freedom, and all of the chi-squared random variables are independent of one another. The exact distribution is too complicated for practical use. Shen and Faraway (2004) suggested using the approximation of Satterthwaite (1941, 1946) and showed that it can be effectively approximated by an ordinary F distribution with degrees of freedom $df_1 = \lambda(p-q)$ and $df_2 = \lambda(n-p)$, where

$$\lambda = \left(\sum_{k=1}^\infty \lambda_k\right)^2 \Big/ \sum_{k=1}^\infty \lambda_k^2 \tag{3}$$

was called the *degrees-of-freedom adjustment factor* by Shen and Faraway (2004).

Model selection is an important issue in regression analysis. Stepwise model selection requires an easy way to calibrate the p value of a predictor in the full model, that is, to test $\beta_j(t) \equiv 0$. This can be done by fitting a reduced model without the jth covariate and using the functional F test statistic

$$F_j = \frac{rss_j - rss}{rss/(n-p)},$$

where rss_j is the residual sum of squares under $\beta_j(t) \equiv 0$. Shen and Faraway (2004) showed that fitting the reduced model is indeed unnecessary, because F_j can be derived from quantities obtained directly from fitting of the full model, that is,

$$F_j = \frac{\|\hat{\beta}_j\|^2}{(rss/(n-p))(\mathbf{X}^T\mathbf{X})_{jj}^{-1}} = \frac{\int \hat{\beta}_j^2(t)\,dt}{(rss/(n-p))(\mathbf{X}^T\mathbf{X})_{jj}^{-1}}, \tag{4}$$

where $(\mathbf{X}^T\mathbf{X})_{jj}^{-1}$ is the jth diagonal element of $(\mathbf{X}^T\mathbf{X})^{-1}$, $\hat{\beta}_j(t)$ is the estimate of $\beta_j(t)$, and rss is the residual sum of squares under the full model. The null distribution of the functional F statistic F_j can be approximated by an ordinary F distribution with degrees of freedom $df_1 = \lambda$ and $df_2 = \lambda(n-p)$, where λ is the degrees-of-freedom adjustment factor defined in (3).

In practice, we do not observe $y_i(t)$ for all t, but only $y_i(t_{ij})$. It is desirable to collect data at fixed time points $t_1, \ldots, t_m$ for easy interpretation and estimation. This occurs in many designed experiments or studies. It is possible that the t_{ij} are different for different i. In such a case, smoothing techniques can be used to get fixed time points (see Faraway 1997). Here we simply assume that the responses are observed on evenly spaced fixed time points, $t_1, \ldots, t_m$. Then (1) becomes

$$y_i(t_j) = \mathbf{x}_i^T\boldsymbol{\beta}(t_j) + \epsilon(t_j) \quad \text{for } i = 1, \ldots, n, j = 1, \ldots, m,$$

and we can do pointwise estimation and regression. We replace the integration with summation and compute $\|\hat{\epsilon}_i\|^2 = \sum_{k=1}^m \hat{\epsilon}_i(t_k)^2/m$ and $\|\hat{\beta}_j\|^2 = \sum_{k=1}^m \hat{\beta}_j(t_k)^2/m$. We can estimate the covariance function $\gamma(s,t)$ by the empirical covariance matrix $\hat{\boldsymbol{\Sigma}} = (\sum_{i=1}^n \hat{\epsilon}_i(t_j)\hat{\epsilon}_i(t_k)/(n-p))_{m\times m}$ and estimate the degrees-of-freedom adjustment factor by

$$\hat{\lambda} = \operatorname{trace}(\hat{\boldsymbol{\Sigma}})^2/\operatorname{trace}(\hat{\boldsymbol{\Sigma}}^2). \tag{5}$$

Large degrees of freedom (say $n - p \geq 30$) are desired for a good estimation of λ.

3. DIAGNOSTICS

Diagnostics are as important for functional regression as for scalar regression. We use residuals to check various assumptions and to identify potential outliers and influential cases.

A straightforward approach is to perform pointwise diagnostics. Pointwise residual plots and normal Q–Q plots are useful for detecting certain violations of assumptions and outliers. However, this approach ignores the fact that the residuals are correlated over time t. In the spirit of the functional F test, we develop diagnostic procedures for functional regression that view each curve as a point in a functional space.

3.1 Studentized Residuals

Let $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ be the hat matrix. The leverage h_{ii} for the ith case is the ith diagonal element of $\mathbf{H}$. For fixed t, the standard results in scalar regression show that $\operatorname{var}(\hat{\epsilon}_i(t)) = (1-h_{ii})\gamma(t,t)$; therefore,

$$E(\|\hat{\epsilon}_i\|^2) = E\int \hat{\epsilon}_i(t)^2\,dt = \int E(\hat{\epsilon}_i(t)^2)\,dt$$
$$= (1-h_{ii})\int \gamma(t,t)\,dt = (1-h_{ii})\Delta^2,$$

where $\Delta^2 = \int \gamma(t,t)\,dt$ is the total variance of the error process. Because $\sum_{i=1}^n (1-h_{ii}) = n-p$, $\hat{\Delta}^2 = rss/(n-p)$ is an unbiased estimate of Δ^2. Thus we formally define (internally) L_2 *studentized residuals* as

$$S_i = \frac{\|\hat{\epsilon}_i\|}{\sqrt{1-h_{ii}}\hat{\Delta}} = \frac{\sqrt{\int \hat{\epsilon}_i^2(t)\,dt}}{\sqrt{1-h_{ii}}\sqrt{rss/(n-p)}}. \tag{6}$$

The L_2 studentized residuals can be used in a similar way as in scalar regression. For example, we can make a residuals versus fitted plot by plotting S_i against $\|\hat{y}_i\|$. Such a plot can be used to check the assumption of equal total variance and to detect potential outliers. As in scalar regression, if the residuals follow the same error process, then S_i and $\|\hat{y}_i\|$ are uncorrelated; thus, we expect to see no relationship between S_i and $\|\hat{y}_i\|$.

To check the assumption of Gaussian process, we study the distribution of the residuals. We have the following important result, which can be proved similarly as theorem 1 of Shen and Faraway (2004).

Theorem 1. If the error process is Gaussian, then $\|\hat{\epsilon}_i\|^2$ is distributed like a linear combination of infinite independent chi-squared random variables with 1 degree of freedom, that is,

$$\|\hat{\epsilon}_i\|^2 \sim (1-h_{ii})\sum_{k=1}^{\infty}\lambda_k\chi_k^2(1),$$

where λ_k are eigenvalues of the covariance function $\gamma(s,t)$ and all of the chi-squared random variables are independent of one another.

This result indicates that $\|\hat{\epsilon}_i\|^2/(1-h_{ii})$ are identically distributed. Using the Satterthwaite approximation, $\|\hat{\epsilon}_i\|^2/(1-h_{ii})$ are approximately distributed as $c\chi^2(\lambda)$, where c is a constant and λ is the adjustment factor defined in (3). Therefore, to check whether the errors are Gaussian, we estimate λ by (5) and make a chi-squared Q–Q plot by plotting S_i^2 against quantiles of a chi-squared distribution with $\hat{\lambda}$ degrees of freedom. If the model is correct and the Gaussian assumption holds, then the points should be close to a straight line.

3.2 Outliers and Jackknife Residuals

As in the context of scalar regression, we define jackknife residuals for functional regression to detect outliers. Suppose that the ith case is a suspected outlier. We delete the ith case from the data and use the remaining $n-1$ cases to fit the linear model. Let $\hat{\boldsymbol{\beta}}_{(i)}(t)$ be the estimate of $\boldsymbol{\beta}(t)$ and let $rss_{(i)}$ be the residual sum of squares, computed without the ith case. Let $\mathbf{X}_{(i)}$ and $\mathbf{Y}_{(i)}(t)$ be the $\mathbf{X}$ matrix and the $\mathbf{Y}(t)$ vector with the ith row deleted. Then $\hat{\boldsymbol{\beta}}_{(i)}(t) = (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}^T\mathbf{Y}_{(i)}(t)$. For the deleted case, compute the fitted curve $\tilde{y}_i(t) = \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_{(i)}(t)$. Because the ith case is not used in estimation, $y_i(t)$ and $\tilde{y}_i(t)$ are independent for fixed t. The variance of $y_i(t)-\tilde{y}_i(t)$ is

$$\operatorname{var}(y_i(t)-\tilde{y}_i(t)) = \gamma(t,t)+\gamma(t,t)\mathbf{x}_i^T\left(\mathbf{X}_{(i)}^T\mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i.$$

If the ith case is not an outlier, then $E(y_i(t)-\tilde{y}_i(t))\equiv 0$. Then

$$\begin{aligned}E\|y_i-\tilde{y}_i\|^2 &= E\int (y_i(t)-\tilde{y}_i(t))^2\,dt\\ &= \int E(y_i(t)-\tilde{y}_i(t))^2\,dt\\ &= \left(1+\mathbf{x}_i^T\left(\mathbf{X}_{(i)}^T\mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i\right)\int\gamma(t,t)\,dt.\end{aligned}$$

We estimate $\Delta^2 = \int\gamma(t,t)\,dt$ by $\hat{\Delta}_{(i)}^2 = rss_{(i)}/(n-p-1)$. Thus we define L_2 jackknife (or externally studentized) residuals as

$$\begin{aligned}J_i &= \frac{\|y_i-\tilde{y}_i\|}{\sqrt{1+\mathbf{x}_i^T(\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\mathbf{x}_i}\hat{\Delta}_{(i)}}\\ &= \frac{\sqrt{\int(y_i(t)-\tilde{y}_i(t))^2\,dt}}{\sqrt{1+\mathbf{x}_i^T(\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\mathbf{x}_i}\sqrt{rss_{(i)}/(n-p-1)}}.\end{aligned}\tag{7}$$

To derive the distribution of J_i, we take an alternative approach and consider the so-called *mean shift outlier model*. Suppose that the ith case is a candidate for an outlier. Assume that the model for all other cases is

$$y_j(t) = \mathbf{x}_j^T\boldsymbol{\beta}(t)+\epsilon_j(t),\qquad j\neq i,$$

but that for case i, the model is

$$y_i(t) = \mathbf{x}_i^T\boldsymbol{\beta}(t)+\delta(t)+\epsilon_i(t).$$

Then testing whether the ith case is an outlier is equivalent to testing $\delta(t)\equiv 0$. We can create a new predictor variable, say u, with $u_i=1$ and $u_j=0$ for $j\neq i$. By (4), the functional F test for $\delta(t)\equiv 0$ is

$$F_i = \frac{\|\hat{\delta}\|^2}{(n-p-1)^{-1}r\tilde{s}s(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})_{ii}^{-1}},\tag{8}$$

where $\hat{\delta}(t)$ is the estimate of $\delta(t)$, $r\tilde{s}s$ is the residual sum of squares, and $\tilde{\mathbf{X}}$ is the model matrix under the mean shift outlier model. It follows from scalar regression that $\hat{\delta}(t) = y_i(t)-\tilde{y}_i(t)$, $r\tilde{s}s = rss_{(i)}$, and $(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})_{ii}^{-1} = (1-h_{ii})^{-1}$ (see, e.g., Sen and Srivastava 1990, pp. 174–175). It is known in scalar regression (see Weisberg 1985, p. 293) that

$$1+\mathbf{x}_i^T\left(\mathbf{X}_{(i)}^T\mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i = (1-h_{ii})^{-1}.\tag{9}$$

Comparing (7) and (8) yields $F_i = J_i^2$. When $\delta(t)\equiv 0$, the statistic F_i, defined in (8), has a functional F distribution according to Shen and Faraway (2004). The next theorem summarizes the results.

Theorem 2. If the ith case is not an outlier, then J_i^2 is distributed like a ratio of two linear combinations of infinite independent chi-squared random variables, that is,

$$J_i^2 \sim \frac{\sum_{k=1}^{\infty}\lambda_k\chi_k^2(1)}{(n-p-1)^{-1}\sum_{k=1}^{\infty}\lambda_k\chi_k^2(n-p-1)},$$

where the λ_k's are eigenvalues of the covariance function $\gamma(s,t)$ and all of chi-squared random variables are independent of one another.

In practice, we use the Satterthwaite approximation and approximate this functional F distribution by an ordinary F distribution with degrees of freedom $df_1 = \lambda$ and $df_2 = \lambda(n-p-1)$, where λ is the adjustment factor defined in (3). We can estimate λ by (5) and preform an F test to formally detect outliers. Because the test is usually done after looking at the results and is applied for all cases, an adjustment of the significance level, such as the Bonferroni method, should be applied.

Furthermore, to avoid fitting the regression model with a case deleted for n times, as in scalar regression, jackknife residuals can be computed directly from studentized residuals and lever-

ages as

$$J_i = S_i \sqrt{\frac{n-p-1}{n-p-S_i^2}}. \tag{10}$$

The proof is given in the Appendix.

3.3 Influential Cases and Cook's Distance

Cook's distance is useful for identifying influential cases in scalar regression. Here we extend the definition to functional regression.

To determine whether the ith case is influential, we can measure the influence by comparing $\hat{\boldsymbol{\beta}}(t)$ to $\hat{\boldsymbol{\beta}}_{(i)}(t)$, the estimates of $\boldsymbol{\beta}(t)$ with and without the ith case. Formally, define L_2 Cook's distance as

$$D_i = \frac{\int (\hat{\boldsymbol{\beta}}_{(i)}(t) - \hat{\boldsymbol{\beta}}(t))^T (\mathbf{X}^T\mathbf{X})(\hat{\boldsymbol{\beta}}_{(i)}(t) - \hat{\boldsymbol{\beta}}(t))\, dt}{p \cdot rss/(n-p)}. \tag{11}$$

Alternatively, if we define $\hat{\mathbf{Y}}(t) = \mathbf{X}\hat{\boldsymbol{\beta}}(t)$ and $\hat{\mathbf{Y}}_{(i)}(t) = \mathbf{X}\hat{\boldsymbol{\beta}}_{(i)}(t)$, then (11) can be rewritten as

$$D_i = \frac{\int (\hat{\mathbf{Y}}_{(i)}(t) - \hat{\mathbf{Y}}(t))^T (\hat{\mathbf{Y}}_{(i)}(t) - \hat{\mathbf{Y}}(t))\, dt}{p \cdot rss/(n-p)}, \tag{12}$$

which measures the distance between $\hat{\mathbf{Y}}(t)$ and $\hat{\mathbf{Y}}_{(i)}(t)$, the fitted responses with and without the ith case. Cases for which D_i are large have substantial influence on $\hat{\boldsymbol{\beta}}(t)$ and on fitted responses, and the deletion of them may result in important changes in conclusions. As in scalar regression, it can be shown (see the App.) that Cook's distance can be computed directly from the studentized residual and leverage as

$$D_i = \frac{1}{p}\frac{h_{ii}}{1-h_{ii}} S_i^2. \tag{13}$$

This formula will save us much computational time because there is no need to fit n models, each with a case deleted. By (13), a highly influential case must have a large leverage or a large studentized residual. An influential case may be (but may not necessarily be) an outlier.

4. AN EXAMPLE

4.1 A Robust Design Experiment

For illustration, we use one of the experiments reported by Nair et al. (2002). An engineering team conducted a robust parameter design experiment to study the effects of seven process assembly parameters (factors A–G) on the audible noise levels of alternators. The experiment used a 2_{IV}^{7-2} design with defining relation $I = CEFG = ABCDF = ABDEG$. For each experimental combination, 43 measurements of sound pressure levels (responses) were recorded at rotating speeds ranging from 1,000 to 2,500 rpm, where the rotating speed was a signal factor. Figure 1 shows the 32 observed (and fitted) response curves. The original data include four additional replications collected at the high levels of all factors, which we do not use here. (For a further description of the data, see Nair et al. 2002.)

For ease of illustration, we first fit a main-effects model; indeed, none of the two-factor interactions is significant. The low and high levels are coded as -1 and $+1$. Table 1 gives the F statistics and p values. The residual sum of squares is 273.09, and the estimated adjustment factor is 4.81. Note that $\hat{\lambda}(n-p) = (4.81)(32-8) = 115.44$. The p value of effect A is computed as the upper tail probability of 3.27 under an ordinary F distribution with degrees of freedom 4.81 and 115.44. Other p values are obtained similarly. The p values show that D and G are very significant and C and A are significant at the 1% margin.

We then simplify the main-effects model by dropping insignificant terms. Table 2 gives the F statistics and p values for a reduced model with main effects A, C, D, and G. As expected, all effects are now significant at the 1% level. The residual sum of squares is 308.01, and the estimated adjustment factor is 5.14. The p values in Table 2 are computed using this new adjustment factor.

We compare the reduced model with the main-effects model by performing a functional F test as follows. The F statistic in (2) is 1.02, with $df_1 = (4.81)(8-5) = 14.43$ and $df_2 = (4.81)(32-8) = 115.44$, yielding a p value of .44. Thus we accept the reduced model with four main effects and proceed to diagnostics.

Figure 2 shows the diagnostic plots for the reduced model. Both the residuals versus fitted plot and chi-squared Q–Q plot suggest a potential outlier. The jackknife residuals plot and Cook's distance plot confirm that case 16 has the largest jackknife residual and Cook's distance. The formal F test with Bonferoni adjustment declares that case 16 is an outlier at the 5% level. The jackknife residual for case 16 is just above the critical value (the horizontal line on the jackknife plot) at the 5% level. Note the big gap between the observed and fitted response curves for case 16 in Figure 1.

We repeat the foregoing analysis without case 16 and still conclude that A, C, D, and G are significant. Table 3 gives the F statistics and p values. The residual sum of squares is 265.95, and the estimated adjustment factor is 6.06. Note that G is still most significant, but C is more significant than D, and A becomes less significant (p value increases from .0076 to .0256). This time, the diagnostic plots do not show apparent patterns calling for attention.

4.2 Simulation Study of Size and Power

Shen and Faraway (2004) used simulation to study the size and power of the functional F test in comparison with the multivariate likelihood ratio and B-spline–based tests under similar conditions to their ergonomics data. Their simulation suggests that the functional F test has a fairly accurate size and could be quite powerful for some types of covariance structures. Note that our data are much rougher than their ergonomics data. It is of interest to investigate how the F test and diagnostics perform for our data.

We simulated response curves as the weighted average of the predicted curves from the reduced model and the main-effects model plus Gaussian errors with mean 0 and covariance $\hat{\boldsymbol{\Sigma}}$, the empirical covariance matrix from the reduced model without case 16. The weight ran from 0 (corresponding to the reduced model) to 1 (corresponding to the main-effects model) in increments of .1. Note that traditional multivariate tests cannot be applied here, because there are 43 dimensions and only 32 cases. For comparison, we also applied B-spline–based tests, with the

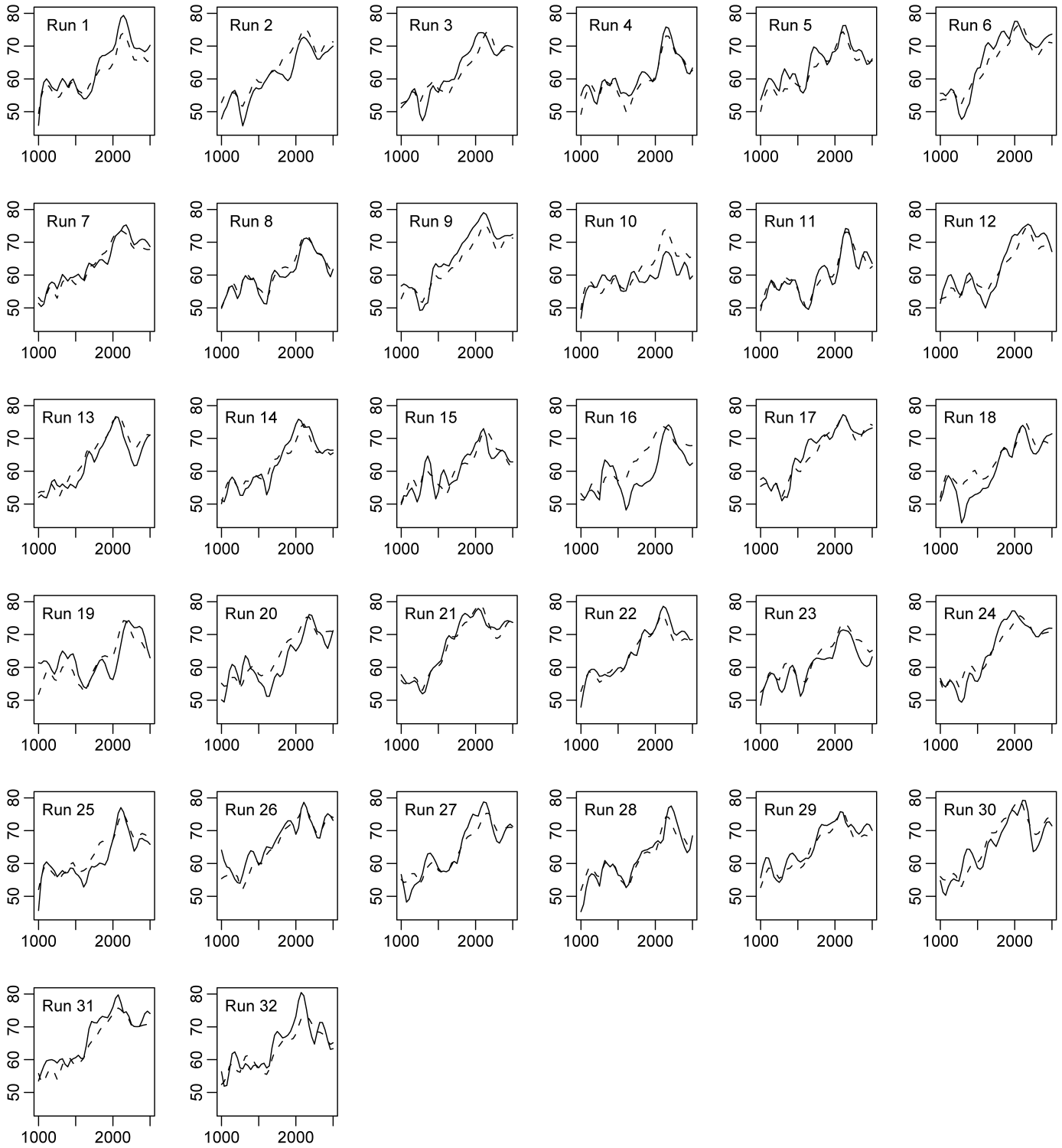

Figure 1. Observed (—-) and Fitted (- - -) Response Curves.

likelihood ratio test applied on the B-spline expansion coefficients. Figure 3 shows the powers (i.e., probabilities of rejecting the reduced model) of the F test and likelihood ratio tests with five and eight B-spline basis functions at significance level .05 over 10,000 repetitions. The F test was comparable to the likelihood ratio test with five B-spline basis functions; the F test had slightly lower power when the weight was $<.6$ and higher power when the weight was $>.6$. The likelihood ratio test with 8 B-spline basis functions was the most powerful among the three tests. We actually computed all likelihood ratio tests with 4–10 B-spline basis functions; of these, the test with 8 B-spline basis functions was the most powerful.

The size of the F test (i.e., the power at weight $= 0$) was .041, below the prespecified significance level of .05, indicating that the F test was conservative here. The reason for this was that the degrees-of-freedom adjustment factor was underestimated in the simulation due to a relatively small sample size. The estimated adjustment factor had an average of 4.90

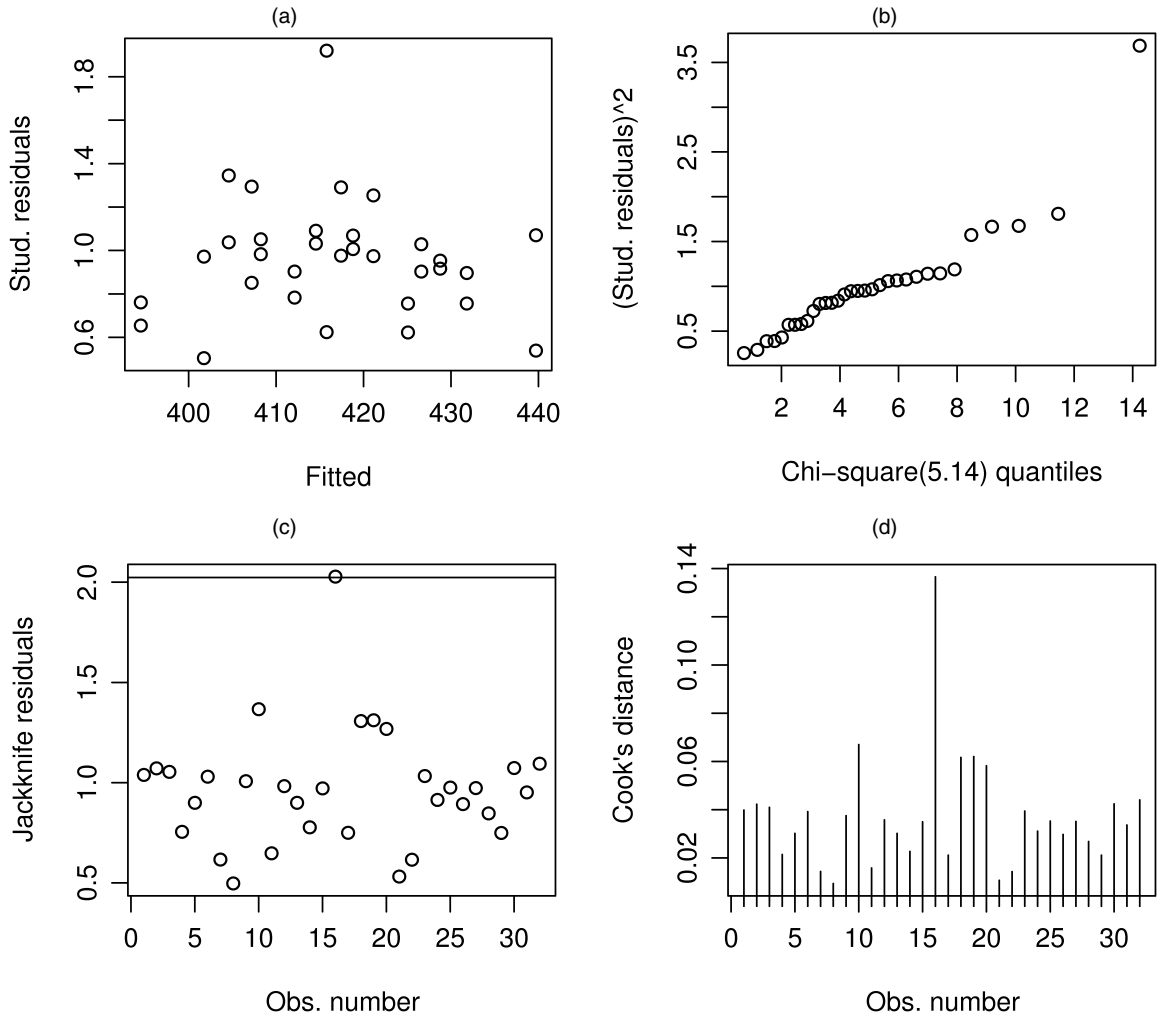

Figure 2. Diagnostic Plots for the Reduced Model: (a) Residuals versus Fitted; (b) Chi-Squared Q–Q Plot; (c) Jackknife Residuals Plots; (d) Cook's Distance Plot.

Table 1. F Statistics and p Values for the Main Effects Model

	A	*B*	*C*	*D*	*E*	*F*	*G*
F value	3.27	1.15	3.61	5.46	.91	1.01	7.81
p value	.0093	.3388	.0051	.0002	.4762	.4117	0

Table 2. F Statistics and p Values for the Reduced Model

	A	*C*	*D*	*G*
F value	3.26	3.60	5.45	7.79
p value	.0076	.0040	.0001	0

Table 3. F Statistics and p Values for the Reduced Model Without Case 16

	A	*C*	*D*	*G*
F value	2.47	5.29	4.29	10.48
p value	.0256	.0001	.0005	0

and standard deviation of .59, whereas the true adjustment factor used in simulation was 6.06. This was not bad, considering that only $32 - 8 = 24$ degrees of freedom were available to estimate a 43×43 covariance matrix. Because we knew the true eigenvalues, we also applied the F test using the true adjustment factor 6.06. With this modification, the F test had a size of .053 and almost the same power as the likelihood ratio test with eight B-spline basis functions.

We conducted another simulation to investigate the performance of diagnostic procedures. For case 16, we simulated the response curve as the weighted average of the observed curve and the predicted curve from the reduced model (without case 16), whereas for all other cases, response curves were simply the predicted curves from the reduced model (without case 16), with added Gaussian errors from the empirical covariance matrix as before. Then we computed jackknife residuals and applied the F test to see whether case 16 was identified as an outlier at significance level .05 over 10,000 simulated datasets. The power curve was similar to that of the F test in Figure 3, consistent with the theory developed in Section 3. The simulated size was .041, and the power at weight $= 1$ was .949. As in the previous simulation, the size was underestimated because the adjustment factor was underestimated (average, 5.01; stan-

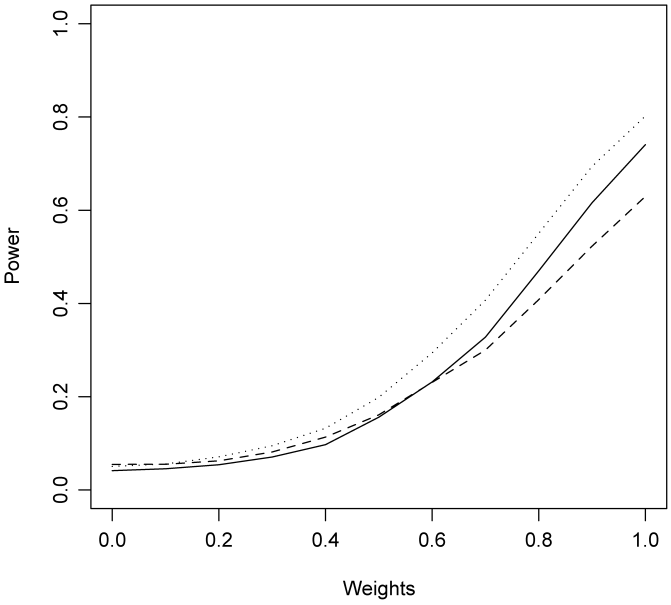

Figure 3. Simulated Power Curves of the F Test (——) and Two B-Spline–Based Tests, One With Five B-Spline Basis Functions (- - - -) and One With Eight B-Spline Basis Functions (·····).

dard deviation, .71). An F test using the true adjustment factor had a size of .051 and a power of .962 when weight = 1.

4.3 Discussion

Faraway (1997) suggested three types of plots for residual analysis. The first type is plots of the estimated eigenfunctions and their associated eigenvalues. These plots show the nature of the unexplained variation in the model and are potentially useful for understanding the error process. Figure 4 shows the first eight estimated eigenfunctions and their associated eigenvalues from the reduced model without case 16. The eight eigenfunctions explain 90% of the variation of the residual functions. The plots indicate that the error process is rather complicated. Determining the dimension of the error process would be difficult. The functional F test does not suffer from this difficulty, which is an advantage over tests based on basis expansion coefficients.

The second type of plot is normal Q–Q plots of the estimated scores of each residual curve. These plots are useful for detecting outliers and for assessing the assumption of Gaussian error process. Typically, we need to examine only a few of these plots associated with the leading eigenvalues. These plots should be examined if the chi-squared Q–Q plot indicates any problems.

The third type is plot of residuals versus fitted for each time point t_j. As in scalar regression, these plots are useful for checking model assumptions and outliers. But it is sometimes difficult to detect outliers when the patterns are not consistent across all time points. For our example, there are 43 plots to be examined, and many (but not all) plots show that case 16 is a potential outlier. Our diagnostic plots clearly show that case 16 is a potential outlier.

5. CONCLUDING REMARKS

Treating each response curve as a point in an L_2 functional space, we have studied residual analysis and diagnostics for

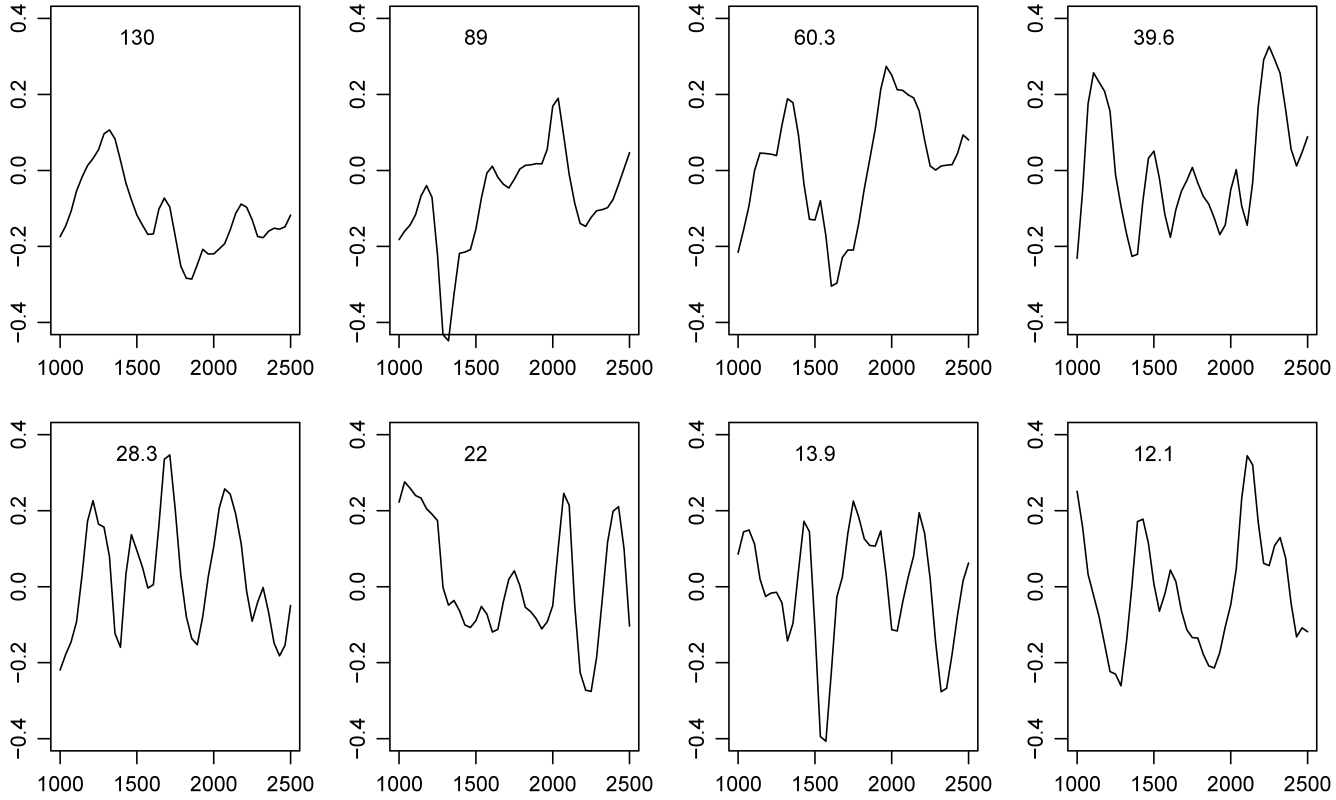

Figure 4. Estimated Eigenfunctions From the Reduced Model Without Case 16. Estimates of the first eight eigenfunctions with the associated eigenvalues marked on the plot.

linear models where the response is a function and the predictors are vectors. Studentized residuals, jackknife residuals, and Cook's distance in the L_2 sense are defined similar to their counterparts in scalar regression. We have discussed their functions in formally detecting outliers and highly influential cases and presented easy computational methods. We gave an example and simulation study to show the effectiveness of these statistics.

When deriving the distribution of our test statistics, we assume that the errors are Gaussian processes. The difficulty of checking multivariate normality is well known in multivariate analysis. Like many other procedures, the chi-squared Q–Q plot provides a necessary and useful check. It can capture nonnormality and outliers in the L_2 sense.

Although the distribution of a linear combination of chi-squared can be computed by numerical integration (as in Imhof 1961) or by simulation, here we use the Satterthwaite approximation. Our various simulations, as well as those of Box (1954), indicate that the Satterthwaite approximation is satisfactory for our purposes. For example, the simulated sizes are near the significance levels when the true adjustment factor is used. Satterthwaite (1941, 1946) and Box (1954) suggested that the Satterthwaite approximation is fairly good when both of the following conditions are met: (a) all coefficients of the chi-squared random variables have the same sign and (b) all chi-squared random variables have the same degrees of freedom. Note that both of these conditions are always met here; therefore, we recommend using the Satterthwaite approximation in practice.

ACKNOWLEDGMENTS

The authors thank the editor, an associate editor, and two referees for their comments.

APPENDIX: PROOFS

Proof of (10)

It is known in scalar regression (see Montgomery 2005, p. 397) that

$$\begin{aligned} y_i(t) - \tilde{y}_i(t) &= (1 - h_{ii})^{-1}(y_i(t) - \hat{y}_i(t)) \\ &= (1 - h_{ii})^{-1}\hat{\epsilon}_i(t). \end{aligned} \tag{A.1}$$

Combining (9) and (A.1) yields

$$J_i^2 = \frac{\|\hat{\epsilon}_i\|^2}{(1 - h_{ii})\hat{\Delta}_{(i)}^2}. \tag{A.2}$$

Let $\hat{\sigma}(t)^2$ and $\hat{\sigma}_{(i)}(t)^2$ be the estimates of variance at time t with and without the ith case. In scalar regression (see Montgomery 2005, p. 398), it is known that

$$(n - p - 1)\hat{\sigma}_{(i)}(t)^2 = (n - p)\hat{\sigma}(t)^2 - (1 - h_{ii})^{-1}\hat{\epsilon}_i(t)^2;$$

then

$$\begin{aligned} rss_{(i)} &= (n - p - 1)\int \hat{\sigma}_{(i)}(t)^2\,dt \\ &= (n - p)\int \hat{\sigma}(t)^2\,dt - (1 - h_{ii})^{-1}\int \hat{\epsilon}_i(t)^2\,dt \\ &= rss - (1 - h_{ii})^{-1}\|\hat{\epsilon}_i\|^2, \end{aligned} \tag{A.3}$$

$$\begin{aligned} \hat{\Delta}_{(i)}^2 &= \frac{rss_{(i)}}{n - p - 1} \\ &= \frac{(n - p)\hat{\Delta}^2 - (1 - h_{ii})^{-1}\|\hat{\epsilon}_i\|^2}{n - p - 1} \\ &= \frac{(n - p - S_i^2)\hat{\Delta}^2}{n - p - 1}. \end{aligned}$$

Finally, combining (A.2), (A.3), and (6) yields

$$J_i^2 = \frac{\|\hat{\epsilon}_i\|^2}{(1 - h_{ii})(n - p - S_i^2)\hat{\Delta}^2/(n - p - 1)} = \frac{(n - p - 1)S_i^2}{n - p - S_i^2}.$$

Proof of (13)

Let $\hat{\sigma}(t)^2$ be the estimate of variance at time t with all cases. Let $d_i(t) = p^{-1}(\hat{\mathbf{Y}}_{(i)}(t) - \hat{\mathbf{Y}}(t))^T(\hat{\mathbf{Y}}_{(i)}(t) - \hat{\mathbf{Y}}(t))\hat{\sigma}(t)^{-2}$ be the Cook's distance at time t. The scalar version of (13) indicates that $d_i(t) = p^{-1}h_{ii}(1 - h_{ii})^{-2}\hat{\epsilon}_i(t)^2\hat{\sigma}(t)^{-2}$. Therefore, $(\hat{\mathbf{Y}}_{(i)}(t) - \hat{\mathbf{Y}}(t))^T(\hat{\mathbf{Y}}_{(i)}(t) - \hat{\mathbf{Y}}(t)) = h_{ii}(1 - h_{ii})^{-2}\hat{\epsilon}_i(t)^2$. Then, by (12) and (6),

$$D_i = \frac{h_{ii}(1 - h_{ii})^{-2}\int \hat{\epsilon}_i(t)^2\,dt}{p \cdot rss/(n - p)} = \frac{1}{p}\frac{h_{ii}}{1 - h_{ii}}S_i^2.$$

[Received September 2005. Revised June 2006.]

REFERENCES

Box, G. E. P. (1954), "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification," *The Annals of Mathematical Statistics*, 25, 290–302.

Eubank, R. L. (2000), "Testing for No Effect by Cosine Series Methods," *Scandinavian Journal of Statistics*, 27, 747–763.

Fan, J., and Lin, S.-K. (1998), "Tests of Significance When Data Are Curves," *Journal of the American Statistical Association*, 93, 1007–1021.

Faraway, J. J. (1997), "Regression Analysis for a Functional Response," *Technometrics*, 39, 254–261.

Imhof, J. P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.

Montgomery, D. C. (2005), *Design and Analysis of Experiments* (6th ed.), New York: Wiley.

Nair, V. N., Taam, W., and Ye, K. Q. (2002), "Analysis of Functional Responses From Robust Design Studies," *Journal of Quality Technology*, 34, 355–370.

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.

Ramsay, J. O., and Silverman, B. W. (2002), *Applied Functional Data Analysis: Methods and Case Studies*, New York: Springer-Verlag.

——— (2005), *Functional Data Analysis* (2nd ed.), New York: Springer-Verlag.

Satterthwaite, F. E. (1941), "Synthesis of Variance," *Psychometrika*, 6, 309–316.

——— (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, 2, 110–114.

Sen, A., and Srivastava, M. (1990), *Regression Analysis: Theory, Methods, and Applications*, New York: Springer-Verlag.

Shen, Q., and Faraway, J. (2004), "An F Test for Linear Models With Functional Responses," *Statistica Sinica*, 14, 1239–1257.

Weisberg, S. (1985), *Applied Linear Regression* (2nd ed.), New York: Wiley.