Statistica Sinica **14**(2004), 1239-1257

AN *F* TEST FOR LINEAR MODELS WITH FUNCTIONAL RESPONSES

Qing Shen and Julian Faraway

Edmunds.com and University of Michigan

Abstract: Linear models where the response is a function, but the predictors are vectors are considered. A functional F test for choosing among two nested functional linear models is developed. Its null distribution is derived and a convenient approximation is presented. A simple way to test individual predictors is presented. The test is applied to some data from Ergonomics and compared to some competing tests. The ability to detect certain types of differences between models is explored. A simulation study is conducted to assess the size and power of the tests.

Key words and phrases: ANOVA model, F test, functional data analysis, power, regression.

1. Introduction

The need to analyze functional observations now arises more often as technology for collecting data at high frequency becomes widespread. Here we are interested in modeling functional response data $y_i(t)$, $i = 1, \ldots n$, $t \in \tau$, where τ is a real interval. We wish to build a regression model to predict this response in terms of a vector of predictors x_i of length p. The model takes the familiar form

$$y_i(t) = x_i^T \beta(t) + \epsilon_i(t).$$

We can estimate $\beta(t)$ by least squares as $\hat{\beta}(t) = (X^T X)^{-1} X^T Y(t)$, where X is the usual $n \times p$ design matrix while $Y(t)$ is a vector of functions of length n formed from the $y_i(t)$'s. Each $\epsilon_i(t)$ is an independent, mean zero, Gaussian stochastic process with a covariance function $r(s,t)$ that can be estimated by

$$\hat{r}(s,t) = \frac{1}{n-p}(Y(s) - X\hat{\beta}(s))^T (Y(t) - X\hat{\beta}(t)).$$

In practice, the functional data will be collected at a finite number of points in τ. The model becomes

$$y_i(t_{ij}) = x_i^T \beta(t_{ij}) + \epsilon_i(t_{ij}),$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, which is a special case of the so called varying-coefficient model in longitudinal analysis with time fixed covariate matrix X. A recent review paper by Wu and Yu (2002) stressed the advantages of a varying-coefficient model, the structural nonparametric regression model, over traditional parametric models and completely nonparametric models. They presented a class of nonparametric smoothing estimation and inference methods for $\beta(t)$ for the general varying-coefficient model with both time varying or time fixed covariates. In contrast to the smoothing techniques they discussed, like local polynomials, splines and basis function approximation, our pointwise least square estimator $\hat{\beta}(t) = (X^T X)^{-1} X^T Y(t)$ is a non-smoothed, unbiased estimator of $\beta(t)$. In order for this estimator to work well, it is desirable that data be collected over a fixed grid points $t_1, \ldots, t_m$ for each subject i to ensure enough data to estimate $\beta(t_j)$ with relatively small variance for each time point t_j. This type of data collection scheme is often called a *regular design*. If the data is collected over a non-regular design, we may first get a smoothed representation of $y_i(t)$, say $y_i^s(t)$. We then obtain an approximation of $y_i(t)$ over common grid points t_j, $j = 1, \ldots, m$, by $y^s(t_j)$, which can be used to estimate $\beta(t_j)$. If the data for each response curve $y_i(t)$ is already quite smooth and plentiful, the choice of smoothing technique would have little impact on the approximation, $y_i^s(t)$. The data for an ergonomics study that we present here is a good example of this type.

Wu and Yu (2002) also pointed out that, although pointwise or simultaneous confidence bands for the coefficient curve $\beta(t)$ have been constructed using either asymptotic approximations or the "resampling-subject" bootstrap method, the hypothesis testing problem of distinguishing a parametric submodel of $\beta(t)$ remains open in the presence of inter-subject covariance. In this article, we tackle the simple, but important hypothesis testing problem of nested linear models for varying-coefficient models with time fixed covariates.

When a regular design or a sufficiently large amount of data is available to approximate the response curve adequately via smoothing, different finite representations of the functional data are possible. One straightforward choice would be to use an equally spaced grid of m points in τ. Another possibility is to use a basis function representation

$$y_i(t) \approx \sum_{j=1}^{m} y_{ij} B_j(t),$$

where $B_j(t)$, $j = 1, \ldots m$, are the basis functions. Cubic B-splines are a common choice although other choices are possible. Either way, we reduce the function $y_i(t)$ to a vector of length m, y_{ij}, for each case i. We should prefer m to be large, particularly when $y_i(t)$ is observed at high frequency with little noise.

In this manner, we reduce the functional data analysis to multivariate data analysis on the y_{ij}. We might hope that the well-established testing techniques of multivariate regression could then be applied — see standard textbooks such as Johnson and Wichern (2002) or Rencher (2002). Of course, a functional interpretation of the results will be necessary, but we might wish to avoid developing any new methodology.

Unfortunately, it is not so simple. For our now (finite) multivariate regression model we have

$$y_{(n\times m)} = X_{(n\times p)}\beta_{(p\times m)} + \epsilon_{(n\times m)},$$

where $\epsilon_{(n\times m)} = (\epsilon_{ij})$ and $\text{Cov}\left((\epsilon_{i1},\ldots\epsilon_{im})^T\right) = \Sigma$. Suppose we compare a smaller model ω which represents a linear subspace of a larger model Ω, where $\dim(\Omega) = p$ and $\dim(\omega) = q$.

The likelihood ratio test statistic is proportional to

$$\log\frac{|\hat{\Sigma}^{\Omega}|}{|\hat{\Sigma}^{\omega}|} = \sum_{j=1}^{m}\log\frac{\lambda_j^{\Omega}}{\lambda_j^{\omega}},$$

where λ_j^{Ω} and λ_j^{ω} are the decreasingly ordered eigenvalues of the empirical covariance matrices $\hat{\Sigma}^{\Omega}$ and $\hat{\Sigma}^{\omega}$, respectively. We can see that terms $\log(\lambda_j^{\Omega}/\lambda_j^{\omega})$ for large j can dominate this statistic. So this test statistic can easily become overwhelmed by unimportant variations represented by the higher order eigenvectors. Ironically, if we are able to observe the data curve on a very fine grid and no smoothing is involved, so that m may be larger than under the basis expansion representation, the power of this test to detect important differences in the models would tend to decrease.

The multivariate tests discussed by the standard textbooks (Johnson and Wichern (2002) or Rencher (2002)) are Wilk's Lambda (which is a function of the likelihood ratio test statistic), the Lawley-Hotelling trace, the Bartlett-Nanda-Pillai trace and Roy's maximum root. Closer examination of these statistics reveals that all suffer from the same defect.

One possible solution is to restrict m, but this choice will be difficult. We prefer a statistic that is not sensitive to the choice of m and scales well as data quality and frequency increases. Faraway (1997) proposed the difference in the integrated residual sums of squares

$$\begin{aligned} rss_{\omega} - rss_{\Omega} &\approx \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_i(t_j) - \hat{y}_{ij}^{\omega})^2 - \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_i(t_j) - \hat{y}_{ij}^{\Omega})^2 \\ &= \frac{n-p}{m}\,\text{trace}(\hat{\Sigma}^{\omega} - \hat{\Sigma}^{\Omega}). \end{aligned}$$

Since the above test statistic is not scale free, it is more natural, by analogy to the usual F tests employed in regression, to use

$$\mathcal{F} = \frac{(rss_\omega - rss_\Omega)/(p-q)}{rss_\Omega/(n-p)} \approx \frac{\text{trace}(\hat{\Sigma}^\omega - \hat{\Sigma}^\Omega)/(p-q)}{\text{trace}(\hat{\Sigma}^\Omega)/(n-p)},$$

where $rss = \sum_{i=1}^n \int_\tau (y_i(t) - \hat{y}_i(t))^2 dt$.

We derive the null distribution of this functional F statistic and show how it can be simply approximated. Our statistic is easily calculated and the p-value is simple to compute. We compare it to alternative test statistics on both a real example from Ergonomics and by simulation.

Our test statistic is more general than one proposed by Box (1954a, 1954b) for one and two way (columnwisely correlated data) ANOVAs. Box's statistic can be used for multivariate data, but was discarded in favor of the well-known tests mentioned above (see Rencher (2002) for a discussion). However, we show that this forgotten statistic can be competitive in terms of balancing between power and agreement with practical sense, i.e., not easily influenced by changes of unimportant variation directions, for both functional and low dimensional multivariate regression analysis.

Ramsay and Silverman (1997) provide a good introduction to functional data analysis. They also propose computing the (pointwise) F statistic at each t but do not investigate in detail how such a statistic can be used for testing. Fan and Lin (1998) developed a basis function type approach for functional ANOVA models, which limits the size of m to satisfy some asymptotic assumption. They gave an adaptively optimal choice of m for realistic applications. Abramovich, Antoniadis, Sapatinas and Vidakovic (2002) used a wavelet representation for functional ANOVA models and used a test developed by Spokoiny (1996) that avoids the dimensionality problem mentioned above, i.e., the problem of the test statistic being dominated by the changes of unimportant variation directions. They applied their test to EEG data which is far rougher than the data in the ergonomics application that motivated the development of our test. The cubic B-spline expansion method is perhaps more appropriate than most wavelet expansions for smooth data applications. We also wish to compare general regression models not just ANOVA models. Eubank (2000) considered tests for a constant mean function using a cosine basis function approach.

Readers should be aware that the term "functional ANOVA" is also used in a different context where a multivariate function is represented by a decomposition in terms of functions of fewer variables. There is also substantial work where functions of the predictors are estimated and tested. We restrict our interest to a functional response in this article.

We develop the theory behind our test statistic in Section 2 and follow it with an application to real data from Ergonomics in Section 3 where we also try out other test statistics. We also perform some simulation based comparisons. Our conclusions are in Section 4.

2. Theory

We are interested in the following functional regression model:

$$y_{(n\times 1)}(t) = X_{(n\times p)}\beta_{(p\times 1)}(t) + \epsilon_{(n\times 1)}(t),$$

where $\epsilon_{(n\times 1)}(t) = (\epsilon_1(t)\dots\epsilon_n(t))^T$ and each $\epsilon_i(t)$ is an independent realization of a Gaussian stochastic process with mean zero and covariance function $r(s,t)$, for $s,t \in \tau$, where τ is a real interval. If $r(s,t)$ is strictly positive definite and $\int_\tau r(t,t)dt < \infty$, the covariance function $r(s,t)$ has an eigen-decomposition

$$r(s,t) = \sum_{i=1}^{\infty} r_i\phi_i(s)\phi_i(t),$$

where $r_1 \geq r_2 \geq \cdots \geq 0$ are the eigenvalues and $\phi_i(t)$, $i = 1,\dots,\infty$, are eigenfunctions satisfying $\int_\tau \phi_i(t)^2 dt = 1$ and $\int_\tau \phi_i(t)\phi_j(t)dt = 0$, $i \neq j$ (Pezzulli and Silverman (1993)). The condition $\int_\tau r(t,t)dt < \infty$ implies that $\sum_{i=1}^{\infty} r_i < \infty$. Usually, the decreasing sequence r_i dies out quickly so that only the first few terms are of any size.

The eigenfunctions $\phi_i(t)$, $i = 1,\dots,\infty$, form a complete sequence in L^2 space, therefore for each error term we have $\epsilon_i(t) = \sum_{j=1}^{\infty} e_{ij}\phi_j(t)$, where $e_{ij} = \int_\tau \epsilon_i(t)\phi_j(t)dt$.

Suppose $\epsilon_i(t)$, $i = 1,\dots,n$, are independent realizations of a Gaussian process with mean zero and covariance function $r(s,t)$. If $r(s,t)$ is continuous on a closed set τ, then it is a standard result in probability theory (Loeve (1977, Corollary 2, p.151)) that the eigenexpansion coefficients e_{ij}, $i = 1,\dots,n$; $j = 1,\dots,\infty$, are independent normal random variables with mean 0 and variance r_j (the expansion of $\epsilon_i(t)$ is just the well-known Karhunen-Loeve expansion of the Gaussian process).

Consider the comparison of two nested linear models, ω and Ω, where $\dim(\Omega) = p$ and $\dim(\omega) = q$. The model ω results from a linear restriction on the parameters of Ω. Without loss of generality, we write the smaller model ω (null hypothesis) as

$$H_0 : Y(t) = X_1\alpha_1(t) + \epsilon(t) \tag{1}$$

and the larger model Ω (alternative hypothesis) as

$$H_1 : Y(t) = X_1\alpha_1(t) + X_2\alpha_2(t) + \epsilon(t),$$

where $\alpha_1(t) = (\beta_1(t), \ldots, \beta_q(t))^T$, $\alpha_2(t) = (\beta_{q+1}(t), \ldots, \beta_p(t))^T$ and ω is obtained from Ω by setting $\beta_i(t) = 0$ for $i = q+1, \ldots, p$.

In this paper, we propose a new test statistic, similar to the ordinary F statistic, as follows:

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

where $rss = \sum_{i=1}^n \int_\tau (y_i(t) - \hat{y}_i(t))^2 dt$.

Theorem 1. *Assume the error process $\epsilon_i(t)$ is Gaussian with continuous covariance function $r(s,t)$ on a closed interval τ, then under* (1), *the test statistic $\mathcal{F}$ is distributed as*

$$\frac{\sum_{i=1}^\infty r_i \chi^2_{(p-q)}/(p-q)}{\sum_{i=1}^\infty r_i \chi^2_{(n-p)}/(n-p)},$$

where r_i is the ith ordered eigenvalue of $r(s,t)$ and the χ^2 random variables are independent.

Proof. Following the discussion in the beginning of this section, for each response curve $y_i(t)$, coefficient function $\beta_i(t)$ and error process $\epsilon_i(t)$, we have $y_i(t) = \sum_{j=1}^\infty y_{ij}\phi_j(t)$, $\beta_i(t) = \sum_{j=1}^\infty b_{ij}\phi_j(t)$ and $\epsilon_i(t) = \sum_{j=1}^\infty e_{ij}\phi_j(t)$, where $y_{ij} = \int_\tau y_i(t)\phi_j(t)dt$, $b_{ij} = \int_\tau \beta_i(t)\phi_j(t)dt$ and $e_{ij} = \int_\tau \epsilon_i(t)\phi_j(t)dt$.

Now we may re-express the model $Y(t) = X\beta(t) + \epsilon(t)$ as:

$$\begin{pmatrix} y_{11} & y_{12} & \cdots \\ \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots \end{pmatrix} \begin{pmatrix} \phi_1(t) \\ \phi_2(t) \\ \vdots \end{pmatrix}$$
$$= \begin{pmatrix} x_{11} & \ldots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \ldots & x_{np} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \ldots \\ \vdots & \vdots & \vdots \\ b_{p1} & b_{p2} & \ldots \end{pmatrix} \begin{pmatrix} \phi_1(t) \\ \phi_2(t) \\ \vdots \end{pmatrix} + \begin{pmatrix} e_{11} & e_{12} & \ldots \\ \vdots & \vdots & \vdots \\ e_{n1} & e_{n2} & \ldots \end{pmatrix} \begin{pmatrix} \phi_1(t) \\ \phi_2(t) \\ \vdots \end{pmatrix}.$$

Let $Y_{.k}$, $\beta_{.k}$ and $e_{.k}$ denote the vectors $(y_{1k}, \ldots, y_{nk})^T$, $(b_{1k}, \ldots, b_{pk})^T$ and $(e_{1k}, \ldots, e_{nk})^T$, respectively. Then the model can be written as $Y_{.k} = X\beta_{.k} + e_{.k}$, for $k = 1, \ldots, \infty$. We also get the predicted value $\hat{y}_i(t) = \sum_{j=1}^\infty \hat{y}_{ij}\phi_j(t)$ and the estimated coefficient $\hat{\beta}_i(t) = \sum_{j=1}^\infty \hat{b}_{ij}\phi_j(t)$, where $\hat{y}_{ij} = \int_\tau \hat{y}_i(t)\phi_j(t)dt$ and $\hat{b}_{ij} = \int_\tau \hat{\beta}_i(t)\phi_j(t)dt$. Let $\hat{Y}_{.k}$ and $\hat{\beta}_{.k}$ denote the vector $(\hat{y}_{1k}, \ldots, \hat{y}_{nk})^T$ and $(\hat{b}_{1k}, \ldots, \hat{b}_{pk})^T$. Since in the L^2 space the expansion coefficients of the basis are unique, it is easy to show that $\hat{\beta}_{.k} = (X^TX)^{-1}X^TY_{.k}$ and $\hat{Y}_{.k} = P_XY_{.k}$, where P_X is the projection matrix $X(X^TX)^{-1}X^T$. The residual sum of squares is

$$rss = \sum_{i=1}^n \int (y_i(t) - \hat{y}_i(t))^2 dt = \sum_{i=1}^n \sum_{k=1}^\infty (y_{ik} - \hat{y}_{ik})^2 = \sum_{k=1}^\infty \sum_{i=1}^n (y_{ik} - \hat{y}_{ik})^2 = \sum_{k=1}^\infty Y_{.k}^T P_X^\perp Y_{.k},$$

where $P_X^\perp = I - P_X$.

Note that, under (1), $Y_{.k} = X_1\beta_{.k} + e_{.k}$; thus we have

$$rss_\omega = \sum_{k=1}^{\infty} Y_{.k}^T P_{X_1}^\perp Y_{.k} = \sum_{k=1}^{\infty} (X_1\beta_{.k} + e_{.k})^T P_{X_1}^\perp (X_1\beta_{.k} + e_{.k}) = \sum_{k=1}^{\infty} e_{.k}^T P_{X_1}^\perp e_{.k},$$

$$rss_\Omega = \sum_{k=1}^{\infty} Y_{.k}^T P_{(X_1,X_2)}^\perp Y_{.k} = \sum_{k=1}^{\infty} (X_1\beta_{.k} + e_{.k})^T P_{(X_1,X_2)}^\perp (X_1\beta_{.k} + e_{.k})$$
$$= \sum_{k=1}^{\infty} e_{.k}^T P_{(X_1,X_2)}^\perp e_{.k}.$$

Therefore,

$$\mathcal{F} = \frac{(rss_\omega - rss_\Omega)/(p-q)}{rss_\Omega/(n-p)} = \frac{\sum_{k=1}^{\infty} e_{.k}^T (P_{X_1}^\perp - P_{(X_1,X_2)}^\perp) e_{.k}/(p-q)}{\sum_{k=1}^{\infty} e_{.k}^T P_{(X_1,X_2)}^\perp e_{.k}/(n-p)}$$
$$= \frac{\sum_{k=1}^{\infty} e_{.k}^T P_{P_{X_1}^\perp X_2} e_{.k}/(p-q)}{\sum_{k=1}^{\infty} e_{.k}^T P_{(X_1,X_2)}^\perp e_{.k}/(n-p)}$$

is distributed as $(\sum_{k=1}^{\infty} r_k \chi^2_{(p-q)}/(p-q))/(\sum_{k=1}^{\infty} r_k \chi^2_{(n-p)}/(n-p))$ since the e_{ij} are independent normal random variables with mean 0 and variance r_j.

Definition 1. (*Functional F distribution*) The null distribution of the $\mathcal{F}$ test statistic is called the functional F distribution with coefficients $\{r_i,\ i = 1, \ldots, \infty\}$ and degrees of freedom $(p-q, n-p)$.

Remark 1. In practice, the response $y(t)$ is often approximated by a vector, the function evaluated on a grid of m points $t_j, j = 1, \ldots, m$. The regression model restricted to these m grid points becomes a standard multivariate multiple regression linear model:

$$y_{(n\times m)} = X_{(n\times p)}\beta_{(p\times m)} + \epsilon_{(n\times m)},$$

where $y = (y_{ij} = y_i(t_j))$, $\beta = (\beta_{ij} = \beta_i(t_j))$, $\epsilon = (\epsilon_{ij} = \epsilon_i(t_j))$, $E(\epsilon_{i.}) = 0_{(m\times 1)}$ and $\mathrm{Cov}\,(\epsilon_{i.}) = \Sigma = (r(t_i,t_j))_{m\times m}$. Let $rss^m = \sum_{i=1}^n \sum_{j=1}^m (1/m)(y_i(t_j) - \hat{y}_i(t_j))^2$. We obtain the approximation of the functional F test statistic as $\mathcal{F}^m = ((rss_\omega^m - rss_\Omega^m)/(p-q))/(rss_\Omega^m/(n-p))$. Using the technique in the proof of Theorem 1 (replacing the covariance function by the covariance matrix, the eigenfunctions by eigenvectors and the infinite sum by a finite sum), it is easy to show that $\mathcal{F}^m$ is distributed as $(\sum_{k=1}^m r_k^m \chi^2_{(p-q)}/(p-q))/(\sum_{k=1}^m r_k^m \chi^2_{(n-p)}/(n-p))$, where r_k^m, $k = 1, \ldots, m$, is the kth ordered eigenvalue of the covariance matrix $\Sigma = (r(t_i,t_j))$ and all the χ^2 random variables are independent of each other. Theorem 1 provides us the exact distribution of the functional F test statistic $\mathcal{F}$, which is the almost sure limit of $\mathcal{F}^m$ as $m \to \infty$.

Theorem 1 and the above remark ensure our functional F test applies to (i) the domain of t is continuous when each response is a function, and (ii) the domain of t is discrete when each response is a vector. Usually, the eigenvalues $r_i, i = 1, \ldots, \infty$ (or $r_i^m, i = 1, \ldots, m$), used in the functional F distribution are unknown, which forms an obstacle to the application of our functional F test procedure.

When $\sum_{i=1}^{\infty} r_i < \infty$, applying the idea of Satterthwaite's (1941) approximation, we can use $c\chi_f^2$ to approximate $\sum_{i=1}^{\infty} r_i \chi_m^2$, where c and f are determined by matching the first two centered moments, as in the two equations

$$cf = \sum_{i=1}^{\infty} r_i m, \quad 2c^2 f = 2\sum_{i=1}^{\infty} r_i^2 m.$$

Therefore,

$$c = \frac{\sum_{i=1}^{\infty} r_i^2}{\sum_{i=1}^{\infty} r_i}, \quad f = \left[\frac{(\sum_{i=1}^{k} r_i)^2}{\sum_{i=1}^{k} r_i^2} m \right],$$

where $[x]$ denotes the closest integer to x. Now we can approximate $(\sum_{i=1}^{\infty} r_i \chi_{(p-q)}^2)$ $/(\sum_{i=1}^{\infty} r_i \chi_{(n-p)}^2)$ by $(c_1 \chi_{f_1}^2)/(c_2 \chi_{f_2}^2)$, where

$$c_1 = c_2 = \frac{\sum_{i=1}^{\infty} r_i^2}{\sum_{i=1}^{\infty} r_i}, \quad f_1 = \left[\frac{(\sum_{i=1}^{\infty} r_i)^2}{\sum_{i=1}^{\infty} r_i^2} (p-q) \right], \quad f_2 = \left[\frac{(\sum_{i=1}^{\infty} r_i)^2}{\sum_{i=1}^{\infty} r_i^2} (n-p) \right].$$

Hence, the functional F distribution $(\sum_{i=1}^{\infty} r_i \chi_{(p-q)}^2/(p-q))/(\sum_{i=1}^{\infty} r_i \chi_{(n-p)}^2/(n-p))$ is approximated by $(\chi_{f_1}^2/f_1)/(\chi_{f_2}^2/f_2)$, which is an ordinary F distribution with degrees of freedom f_1 and f_2.

Definition 2. (*Degrees-of-freedom-adjustment-factor*) The value $(\sum_{i=1}^{\infty} r_i)^2/\sum_{i=1}^{\infty} r_i^2$ is called the degrees-of-freedom-adjustment-factor, where r_i is the ith eigenvalue of the covariance function $r(s,t)$.

In practice, when we only observe $y_i(t)$ over the grid points $t_j, j = 1, \ldots, m$, the adjustment factor $(\sum_{i=1}^{\infty} r_i)^2/\sum_{i=1}^{\infty} r_i^2$ can be estimated by trace$(E)^2/$trace(E^2), where $E = \hat{\Sigma}^{\Omega}$ is the empirical covariance matrix computed from the full model Ω. The proof of consistency of this estimator can be found in Dauxois, Pousse and Romain (1982). One may worry that a very large grid size m is needed to make this estimator work. Actually, from Remark 1 of Theorem 1 we see that for each fixed grid size m, the test statistic $\mathcal{F}$ approximated on these grid points becomes $\mathcal{F}^m$, which is exactly distributed as a functional F distribution with coefficients $(r_i^m, i = 1, \ldots m)$ where r_i^m is the ith ordered eigenvalue of the covariance matrix $\Sigma = (r(t_i, t_j))$. So as long as we can approximate r_i^m well using the ith ordered eigenvalue of the empirical covariance matrix E, it is fine to use this approximate test with true degrees-of-freedom-adjustment-factor $(\sum_{i=1}^{m} r_i^m)^2/\sum_{i=1}^{m}(r_i^m)^2$. To get a good estimation of r_i^m, it is better that the degree of freedom $n-p$ used in the estimation be sufficiently large, say no less than 30.

The idea of the functional F distribution and its Satterthwaite's approximation are not completely new in the literature. Box (1954b) derived the null distribution of the F statistic in the two-way ANOVA (colomnwisely correlated data) as a special case of our functional F distribution and Satterthwaite's approximation of the null distribution is also recommended. (Actually, the distribution of the between columns test criteria in Box (1954b) can be easily derived from Remark 1 with $p = 1$ and $q = 0$.) Interested readers may refer to Table I and II in Box (1954a) for a simulation study of the accuracy of Satterthwaite's approximation of linear combinations of independent chi-squared random variables and their ratio. Box (1954a) showed that the Satterthwaite's approximation is fairly accurate over a wide range of linear combination coefficients and degrees of freedom of the chi-squared random variables, and may be usefully employed to supplement the accurate (but less suggestive) exact methods. In the next section we also use simulation to show that the approximation of the functional F distribution produces a fairly accurate size for the test.

After fitting the full model

$$H_1 : Y(t) = X\beta(t) + \epsilon(t), \tag{2}$$

we may wonder how significant each covariate is in predicting $Y(t)$, i.e., we want to test

$$H_{0i} : \beta_i(t) = 0, \quad i = 1, \ldots, p, \tag{3}$$

against the full model. For each i, first we can fit the null model, which is the full model dropping the ith covariate. Then we may use the functional F test statistic

$$\mathcal{F}_i = \frac{rss_{0i} - rss_1}{rss_1/(n-p)}$$

to make a decision whether or not to accept the null model. It can be very tedious to do this for all $i = 1, \ldots, p$, since we need to fit a different null model each time. Fortunately the following theorem provides us an easy way of computing the test statistics $\mathcal{F}_i$.

Theorem 2. *For $i = 1, \ldots, p$, $\mathcal{F}_i = \frac{(n-p)\int \hat{\beta}_i^2(t)dt}{rss_1\,(X^TX)_{ii}^{-1}}$.*

Proof. Let X_{0i} and X denote the model matrices for H_{0i} and H_1, respectively. Following the notation in the proof of Theorem 1, we have

$$\frac{rss_{0i} - rss_1}{rss_1} = \frac{\sum_{k=1}^{\infty} Y_{.k}^T P_{X_{0i}}^{\perp} Y_{.k} - \sum_{k=1}^{\infty} Y_{.k}^T P_X^{\perp} Y_{.k}}{rss_1}.$$

It is known in scalar type regression theory (Weisberg (1985)) that

$$Y_{.k}^T P_{X_{0i}}^{\perp} Y_{.k} - Y_{.k}^T P_X^{\perp} Y_{.k} = \frac{\hat{b}_{ik}^2}{(X^T X)_{ii}^{-1}}.$$

Therefore,

$$\mathcal{F}_i = \frac{rss_{0i} - rss_1}{rss_1/(n-p)} = \frac{(n-p)\sum_{k=1}^{\infty} \hat{b}_{ik}^2}{rss_1 (X^T X)_{ii}^{-1}} = \frac{(n-p)\int \hat{\beta}_i^2(t)dt}{rss_1 (X^T X)_{ii}^{-1}}.$$

According to Theorem 2, the functional F test statistic $\mathcal{F}_i$ only involves quantities that can be calculated directly from fitting the full model (2). Thus we reduce the complexity of computation from fitting $p+1$ models to fitting one model only. We reject (3) if $\mathcal{F}_i > \mathcal{F}_{(r;(1,n-p))}^{(1-\alpha)}$, where $\mathcal{F}_{(r;(1,n-p))}^{(1-\alpha)}$ denotes the $1-\alpha$ percentile of the functional F distribution with degrees of freedom 1 and $n-p$. Or, using the approximation of the null distribution, we reject (3) if $\mathcal{F}_i > F^{(1-\alpha)}(f_1, f_2)$, where $f_1 = [\text{trace}(E)^2/\text{trace}(E^2)], f_2 = [(n-p)\,\text{trace}(E)^2/\text{trace}(E^2)]$, and E is the empirical covariance matrix of the error process obtained from (2).

3. Example

3.1. The ergonomics data

As part of a project to predict the motion of drivers of automobiles, researchers at the Center for Ergonomics at the University of Michigan collected data on the motion of a single subject to 20 locations within a test car. Amongst other measures, the angle formed at the right elbow between the upper and lower arm was measured using motion capture equipment. For each reach, there were three replicates. The locations were spread around the glove compartment, the gear shift, the central instrument panel and an overhead panel.

The data recorded for each motion vary in time length. However, since the objective of this study is only to model the shape of the motion, not the speed at which it occurred, for each response curve $y(t)$ we rescaled t to vary over [0,1]. For a given motion, $y(t)$ is observed on an equally spaced grid of points, but the number of such points varies from observation to observation. For each motion, the data was smoothly interpolated and expressed on a grid of 20 points. The right elbow angle curves for one individual are shown in Figure 1. One of the motions to the left rear shifter location is clearly wrong since the subject changed his mind about the target location in mid-reach, so we discard this observation in the rest of the analysis. More details may be found in Faraway (1997). The data is available from the website of the second author.

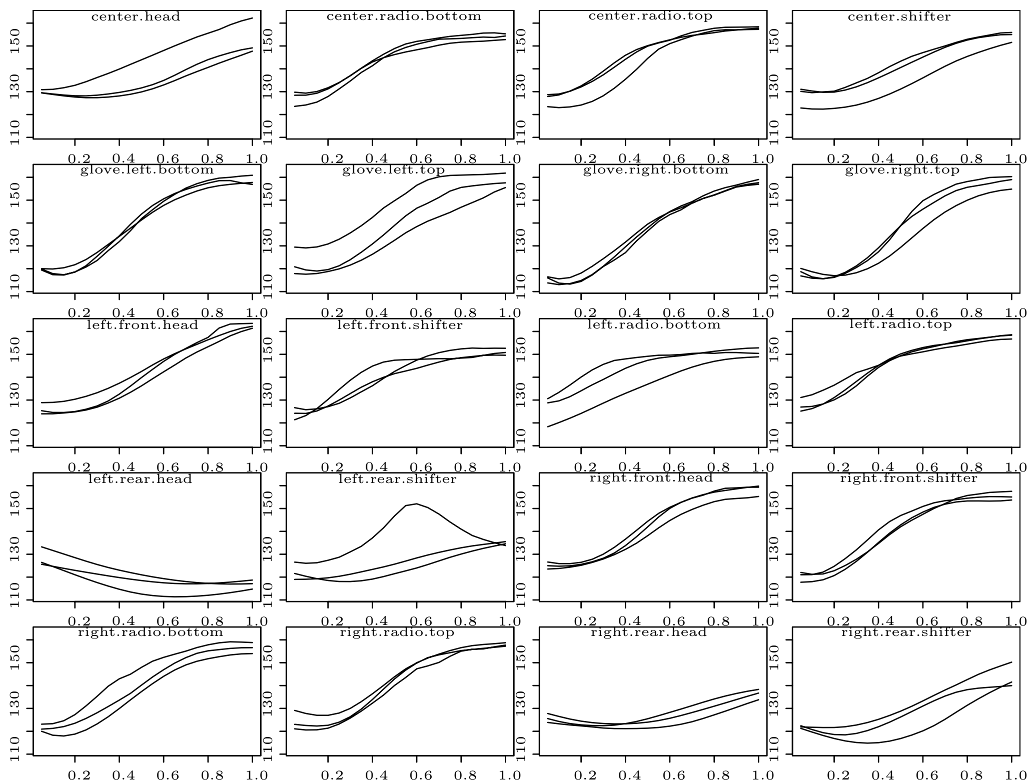

Figure 1. Smoothed right elbow angle curves.

3.2. Comparison of four testing procedures

The main purpose of this experiment is to find a model for predicting the motion given the coordinates (c_x, c_y, c_z) of the target, where x represents the left to right direction, y represents the close to far direction, and z represents the down to up direction. A linear model

$$y(t) = \beta_0(t) + c_x\beta_x(t) + c_y\beta_y(t) + c_z\beta_z(t) + \epsilon(t)$$

and a quadratic model

$$y(t) = \beta_0(t) + c_x\beta_x(t) + c_y\beta_y(t) + c_z\beta_z(t) + c_xc_y\beta_{xy}(t) + c_yc_z\beta_{yz}(t) + c_xc_z\beta_{xz}(t) + c_x^2\beta_{x^2}(t) + c_y^2\beta_{y^2}(t) + c_z^2\beta_{z^2}(t) + \epsilon(t)$$

were fit to the data and compared with the one-way ANOVA model

$$y(t) = \beta_k(t) + \epsilon(t),$$

where $k = 1, \ldots, 20$ indicates the target. Comparing the linear or quadratic models to the ANOVA model represents a lack of fit test. Figure 2 presents the fitted curves of the ANOVA model, the linear model and the quadratic model. It is clear that the quadratic fit is much better than the linear fit, which becomes worse approaching the end of each reach. To make formal inferential comparisons, we use four different testing procedures: the bootstrap methods as described in Faraway (1997), our functional F test, the traditional multivariate log likelihood ratio test on raw data and a test based on a B-spline basis function representation. In the B-spline method, we represent each curve as a linear combination of eight B-splines and then perform the usual multivariate test on the coefficients of this representation. The p-values for these tests are shown in Table 1.

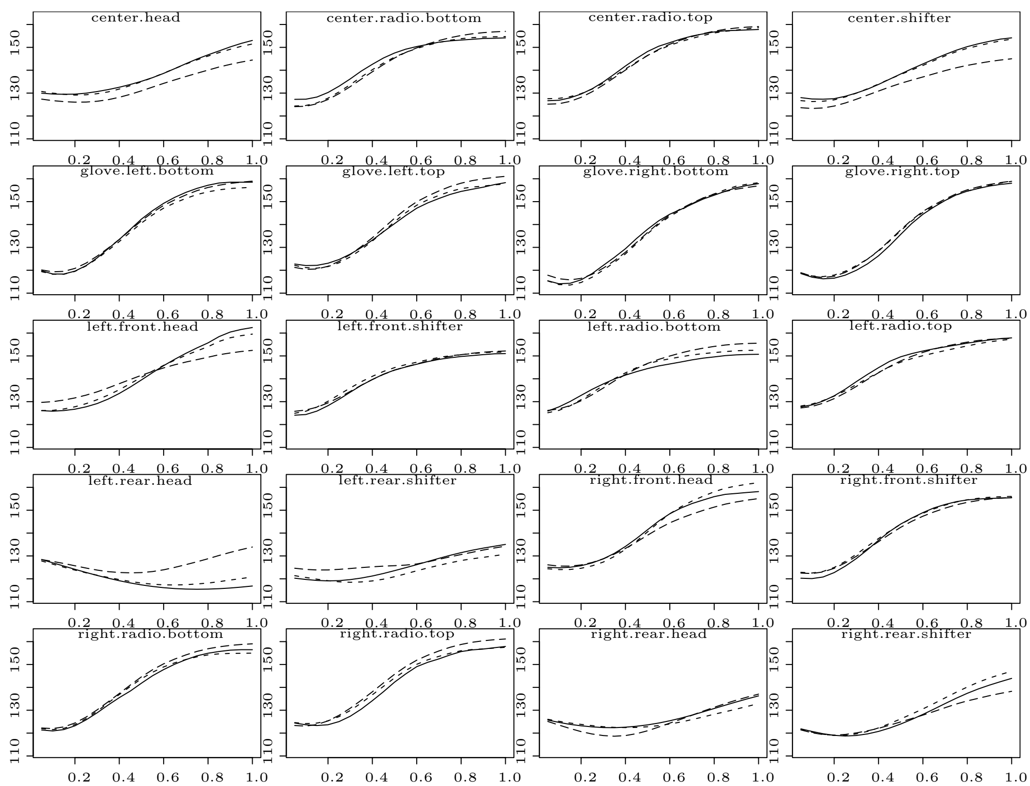

Figure 2. Predicted right elbow angle curves from the ANOVA model (solid line), the quadratic model (dotted line) and the linear model (dashed line).

Table 1. Comparison of testing procedures. P-values shown.

	Bootstrap	Functional F	LRT	B-Spline
Linear vs. ANOVA	0.00	0.00	0.00	0.00
Quadratic vs. ANOVA	0.26	0.53	0.008	0.018

We see that the linear model is clearly rejected as an acceptable fit to the data by all four procedures but there is a difference of opinion regarding the quadratic model. Should we accept the quadratic model or not? In Figure 3, we show the first four eigenvectors and eigenvalues of the residuals of the ANOVA, quadratic and linear models. We find that the shape of the four eigenvectors are very similar for the ANOVA and quadratic models while the linear model has a distinguished first eigenvector and eigenvalue from the other two models. It is easy to explain why the linear model has been rejected by all tests. The quadratic model has been rejected by the traditional multivariate log likelihood ratio test, since the second, third, fourth and and even some smaller eigenvalues of the quadratic model are all about two times those of the ANOVA model. But, since the dominating first eigenvalue of the quadratic model is only 10 percent higher than that of the ANOVA model, the functional F test statistic, which is equivalent to the ratio of the sum of the eigenvalues from the quadratic and ANOVA models, would not be large enough to reject the quadratic model. Here we would prefer to accept the quadratic model because it captures the size of the dominating variation. After all, a model should capture the main features of the data and not be rejected for unimportant reasons.

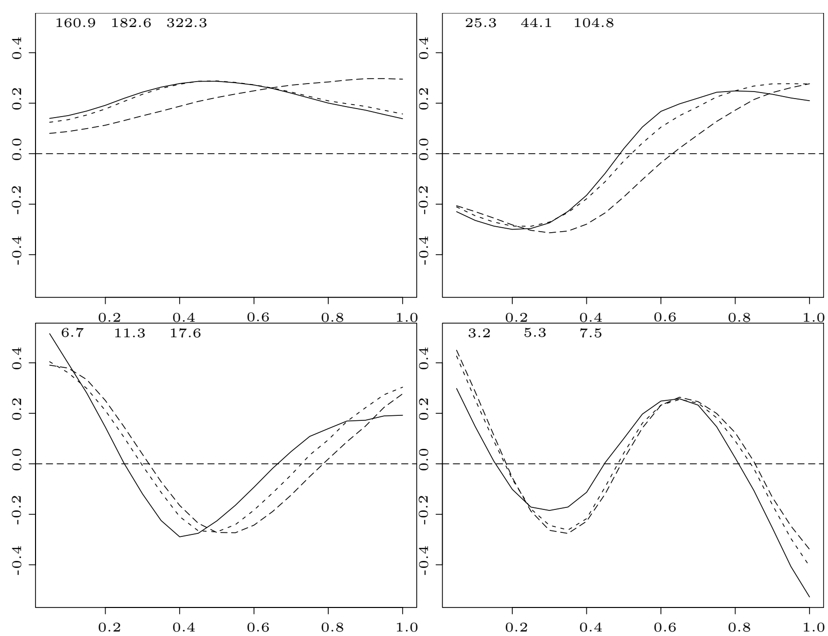

Figure 3. Residual principal components from the ANOVA model (solid line), the quadratic model (dotted line) and the linear model (dashed line). In the top line of each graph, the three numbers from left to right are the eigenvalues from the ANOVA, quadratic and linear models, respectively.

Ramsay and Silverman (1997) suggested a plot of the pointwise F statistics over the grid points. This is shown in Figure 4 together with the critical values at 0.05 level (Bonferroni corrected or not). We see that for the comparison of the linear v.s. the ANOVA model, almost half of the F test statistics are greater than the critical values, which naturally leads to the rejection of the linear model. However, for the comparison of the quadratic v.s. the ANOVA model, most of the test statistics are below the uncorrected critical value and all of them are below the Bonferroni corrected critical value, which also suggests there is not enough evidence to reject the quadratic model.

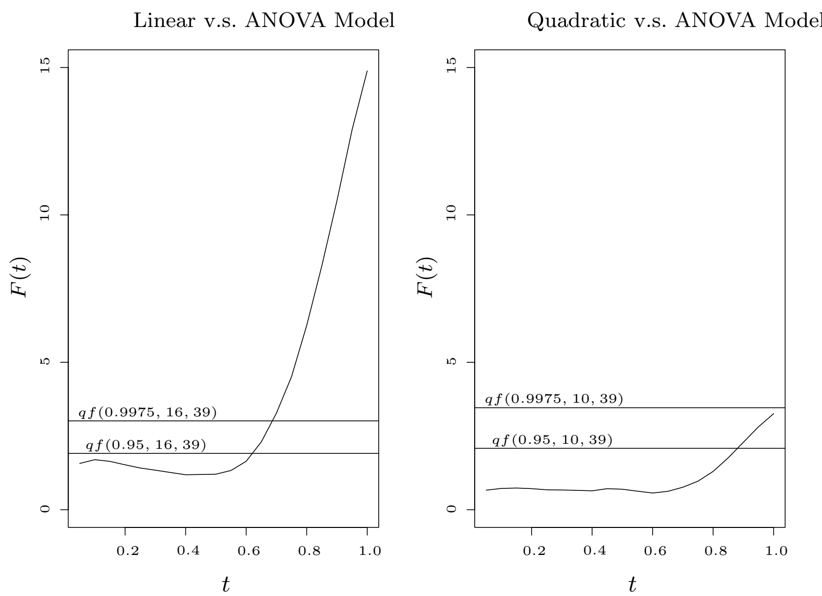

Figure 4. Pointwise F statistics with critical values at 0.05 level (Bonferroni corrected or not).

3.3. Simulation study of size and power

To study the size and power of the functional F test together with the multivariate log likelihood ratio test and the B-spline based test, we ran simulation studies under similar conditions to our ergonomics data. Response curves were simulated using the weighted average of the predicted curves from the quadratic model and the ANOVA model plus the simulated error process with two types of covariance structure. The weight runs from 0 (which produces the quadratic fit) to 1 (which produces the ANOVA fit) in increments of 0.05. At each weight we simulated 1000 sets of response curves and applied the functional F test, the multivariate likelihood ratio test and the B-spline test, respectively, at significance level 0.05, to compare the quadratic model and the ANOVA model. The resulting probability of rejecting the quadratic model is plotted in Figure 5. The top plot shows the result of using the error process with the covariance matrix set to be the empirical covariance matrix from the ANOVA model of the observed data, and the bottom plot shows the result with a covariance matrix of the form $(10 \times 0.8^{|i-j|})$. When the weight is zero, the quadratic model is the true model so the probability of rejection is the size of the test. For all three tests the nominal significance level is 0.05 but, because of various approximations, we expect some error. For the first error structure, the estimated size for the functional F test, the multivariate log likelihood ratio test and the B-spline test are $0.056, 0.087, 0.071$, respectively. For the second error structure, the estimated size for the three tests are $0.064, 0.087, 0.066$, respectively. We see that under both conditions the functional F test has the most accurate size.

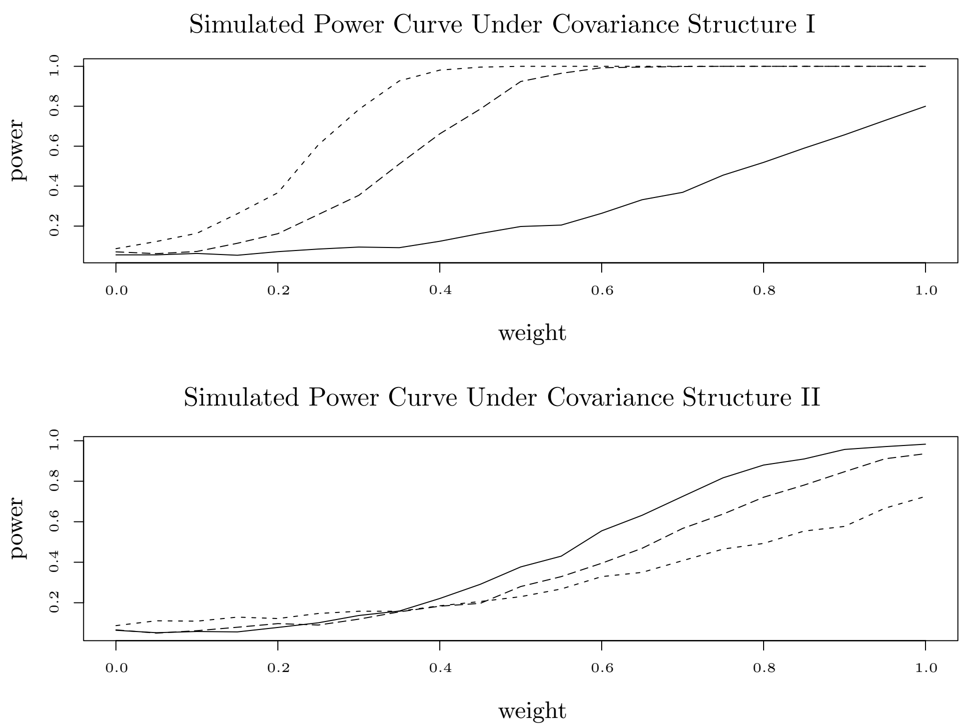

Figure 5. Simulated power curves from the functional F test (solid line), the traditional multivariate log likelihood ratio (Wilk's Lambda) test (dotted line) and the B-spline multivariate test (dashed line). Covariance structure I is the empirical covariance matrix estimated from the ANOVA model; covariance structure II $= (10 \times 0.8^{|i-j|})$.

When the weight is greater than 0, the true model is no longer the quadratic model, so the probability of rejecting the quadratic model becomes the power of the test. For the first error structure, the multivariate log likelihood ratio test is always the most powerful one, the B-spline test is the second and the functional F test is the least powerful. So the test result of non-rejection of the quadratic model based on the functional F test may be caused by lack of power. For the second error structure, when weight is less than 0.35, all three tests have similar power; when the weight is greater than 0.35, the functional F test is the most powerful, the B-spline test is the second and the multivariate log likelihood ratio test is the least powerful.

We see that the covariance structure of the error process has a big influence of the power of the tests. For the first error structure, the ordered eigenvalues of the covariance matrix are $(182.6, 44.1, 11.3, 6.3, \ldots)$. For the second error structure the ordered eigenvalues of the covariance matrix are $(72.3, 43.2, 24.5, 14.7, \ldots)$, which decrease at a much slower rate than the eigenvalues of the first error structure. Actually, the size of the degrees-of-freedom-adjustment-factor $(\sum_{i=1}^{\infty} r_i)^2 / \sum_{i=1}^{\infty} r_i^2$, where r_i is the ith ordered eigenvalue of the error covariance matrix, is determined not by the actual size of the eigenvalues but by the decreasing rate of the eigenvalue sequence. The faster the eigenvalues decrease the smaller the degrees-of-freedom-adjustment-factor would be. A conjecture we have based on this fact is that the faster the eigenvalues decrease the smaller the power of the functional F test tends to be. Our simulation studies have shown that, when error covariance structure is proportional to $(q^{|i-j|})$ where $q \leq 0.9$ and the grid size is 20, the functional F test would be the most powerful test among the three tests we have considered. When the grid size is larger than 20, q can be even larger and still the functional F will be the most powerful among the three tests.

In Figure 6, we plot one set of the response curves under the second type of error covariance structure when the weight is 0. We see that the curves are much rougher than our observed data. For such within-curve covariance structure, we would clearly recommend the functional F test since our simulation studies have shown that the functional F test has the highest power among all the four tests considered. In contrast, if the error covariance structure is like the first type, where the eigenvalues of the covariance matrix decrease quickly, the likelihood ratio test is more powerful than the function F test. But, as discussed in the introduction, the likelihood ratio test statistic may be influenced by the unimportant directions of variation (smaller eigenvalues) and so produce statistically significant, but practically less meaningful, test results.

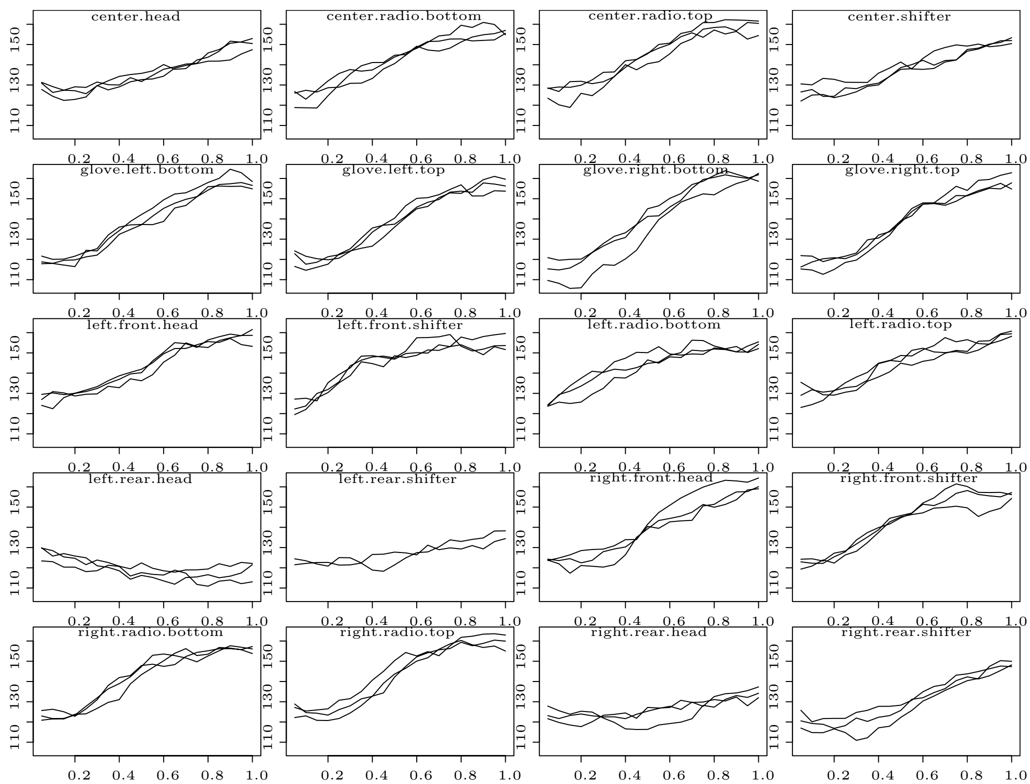

Figure 6. Simulated right elbow angle curves from the quadratic model with covariance structure II=$(10 \times 0.8^{|i-j|})$.

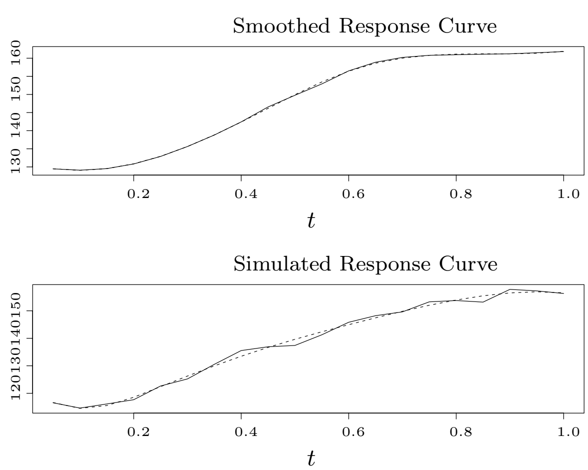

Figure 7. Right elbow response curve (solid line) v.s. B-spline fitting curve (dotted line). The top panel is under covariance structure I and the bottom panel is under covariance structure II.

The B-spline test is an improvement over the traditional multivariate tests. However it may not be possible to avoid the influence of unimportant directions, depending on how many basis functions are chosen to represent the raw data. In Figure 7, we plot the raw data v.s. the 8-basis B-spline expansion curve under both error covariance structures. We see that for the first error structure, the 8-basis B-spline expansion curve represents the raw data perfectly, while for the second error structure it is too smooth to capture all the features of the raw data. In fact, we have to employ a very large number of bases to fit the raw data closely in the second situation, which would make the B-spline method not much different from the traditional multivariate tests. Of course, we could use fewer basis functions solely for the purpose of making the test, but it is difficult to know how many. Our functional F test avoids this dilemma.

4. Conclusion

The advantages of the functional F test are that it is simply computed and does not require the user to be careful about the grid size or the number of basis functions required. The null distribution of the test statistics is easily approximated and approximate p-values are simple to compute. Depending on the data, it may or may not be the most powerful test for comparing two nested functional linear models, but we claim that it examines important rather than trivial differences between models. Furthermore, we see that our test can compare favorably with other well known tests such as Wilk's Lambda, even for non-functional multivariate data.

Acknowledgements

The authors thank the Editor, an associate editor, and the referees for their valuable comments and suggestions.

References

Abramovich, F., Antoniadis, A., Sapatinas, T. and Vidakovic, B. (2002). Optimal testing in functional analysis of variance models. Georgia Institute of Technology ISYE Statistics Technical Report.

Box, G. E. P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems I. Effects of inequality of variance in the one-way classification. *Ann. Math. Statist.* **25**, 290-302.

Box, G. E. P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Ann. Math. Statist.* **25**, 484-498.

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function. *J. Multivariate Anal.* **12**, 136-154.

Eubank, R. L. (2000). Testing for no effect by cosine series methods. *Scand. J. Statist.* **27**, 747-763.

Fan, J. and Lin, S.-K. (1998). Tests of significance when data are curves. *J. Amer. Statist. Assoc.* **93**, 1007-1021.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics* **39**, 254-261.

Johnson, R. and Wichern, D. (2002). *Applied Multivariate Statistical Analysis*. 5th edition. Prentice Hall, New Jersey.

Loeve, M. (1977). *Probability Theory II*. 4th edition. Springer-Verlag, New York.

Pezzulli, S. and Silverman, B. W. (1993). Some properties of smoothed principal components analysis for functional data. *Comput. Statist.* **8**, 1-16.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer, New York.

Rencher, A. C. (2002). *Methods of Multivariate Analysis*. 2nd edition. Wiley, New York.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika* **6**, 309-316.

Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24**, 2477-2498.

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

Wu, C. O. and Yu, K. F. (2002). Nonparametric varying-coefficient models for the analysis of longitudinal data. *Internat. Statist. Rev.* **70**, 373-393.

Edmunds.com, Inc, 2401 Colorado Ave., Suite 250, Santa Monica, CA 90404, U.S.A.

E-mail: shenqing26@hotmail.com

Department of Statistics, University of Michigan, Ann Arbor, MI 48109–1092, U.S.A.

E-mail: faraway@umich.edu

(Received May 2003; accepted January 2004)