Statistica Sinica **21** (2011), 1431-1451

STATISTICAL INFERENCES FOR LINEAR MODELS WITH FUNCTIONAL RESPONSES

Jin-Ting Zhang

National University of Singapore

Abstract: With modern technology development, functional responses are observed frequently in fields such as biology, meteorology, and ergonomics, among others. Consider statistical inferences for functional linear models in which the response functions depend on a few time-independent covariates, but the covariate effects are functions of time. Of interest is a test of a general linear hypothesis about the covariate effects. Existing test procedures include the L^2-norm based test proposed by Zhang and Chen (2007) and the F-type test proposed by Shen and Faraway (2004), among others. However, the asymptotic powers of these testing procedures have not been studied, and the null distributions of the test statistics are approximated using a naive method. In this paper, we investigate the F-type test for the general linear hypothesis and derive its asymptotic power. We show that the F-type test is root-n consistent. In addition, we propose a bias-reduced method to approximate the null distribution of the F-type test. A simulation study demonstrates that the bias-reduced method and the naive method perform similarly when the data are highly or moderately correlated, but the former outperforms the latter significantly when the data are nearly uncorrelated. The F-type test with the bias-reduced method is illustrated via applications to a functional data set collected in ergonomics.

Key words and phrases: Functional data, functional hypothesis test, F-type test, Gaussian process, root-n consistency, χ^2-type mixtures, χ^2-approximation.

1. Introduction

Functional data consist of functions that are often smooth and usually corrupted with noise. With modern technology development, such functional data are observed frequently in fields such as biology, meteorology, and ergonomics, among others; see Besse and Ramsay (1986), Ramsay (1995), and Ramsay and Dalzell (1991) among others for good examples and analyses. Comprehensive surveys about functional data analysis (FDA) can be found in Ramsay and Silverman (1997, 2002).

This paper is motivated by the ergonomics data downloaded from the website of the second author of Shen and Faraway (2004). To study the motion of drivers of automobiles, the researchers at the Center for Ergonomics in the University of

Michigan collected data on the motion of a single subject to 20 locations within a test car. Among others, the researchers measured 3 times the angle formed at the right elbow between the upper and lower arms. The data recorded for each motion were observed on an equally spaced grid of points over a period of time, that was rescaled to $[0, 1]$ for convenience, but the number of such time points varied from observation to observation. See Faraway (1997) and Shen and Faraway (2004) for a detailed description of the data and Figure 1 of Shen and Faraway (2004) for smoothed right elbow curves. To find a model for predicting the right elbow angle curve $y(t), t \in [0, 1]$ given the coordinates (a, b, c) of the target, where a represents the "left to right" direction, b represents the "close to far" direction, and c represents the "down to up" direction, Shen and Faraway (2004) compared a linear model, a quadratic model and a one-way ANOVA model and found the quadratic model adequate to fit the data. The quadratic model they considered may be written as

$$\begin{aligned} y_i(t) = \beta_0(t) + a_i\beta_1(t) + b_i\beta_2(t) + c_i\beta_3(t) + a_i^2\beta_4(t) \\ +b_i^2\beta_5(t) + c_i^2\beta_6(t) + a_ib_i\beta_7(t) + a_ic_i\beta_8(t) \\ +b_ic_i\beta_9(t) + v_i(t), i = 1, \ldots, 60, \end{aligned} \tag{1.1}$$

where $y_i(t)$ and $v_i(t)$ denote the i-th response and location-effect curves over time, respectively, (a_i, b_i, c_i) denotes the coordinates of the target associated with the i-th angle curve, and $\beta_r(t), r = 0, 1, \ldots, 9$, are unknown coefficient functions. Some questions then arise naturally. Is each of the coefficient functions significant? Can the quadratic model (1.1) be further reduced? These two and other similar questions can be answered easily after we study the general linear hypothesis testing problem given below.

Consider a functional linear model (FLM) with functional responses:

$$y_i(t) = \boldsymbol{x}_i^T\boldsymbol{\beta}(t) + v_i(t),\ v_i(t) \stackrel{i.i.d.}{\sim} \mathrm{GP}(0, \gamma), t \in T,\ i = 1, \ldots, n, \tag{1.2}$$

where $y_i(t),\ i = 1, \ldots, n$, are the response functions, $\boldsymbol{x}_i,\ i = 1, \ldots, n$, are the time-independent $(p+1)$-dimensional covariates, $\boldsymbol{\beta}(t) = [\beta_0(t), \beta_1(t), \ldots, \beta_p(t)]^T$ is the $(p + 1)$-dimensional vector of coefficient functions, $v_i(t),\ i = 1, \ldots, n$, are the subject-effect curves, $T = [a, b], -\infty < a < b < \infty$, is the support of the design time points, and here and throughout $\mathrm{GP}(\eta, \gamma)$ denotes a Gaussian process with mean function $\eta(t)$ and covariance function $\gamma(s, t)$. It is clear that the FLM (1.1) is a special case of (1.2). Based on the FLM (1.2), a general linear hypothesis testing (GLHT) problem is

$$H_0 : \boldsymbol{C}\boldsymbol{\beta}(t) = \boldsymbol{c}(t),\ \text{vs } H_1 : \boldsymbol{C}\boldsymbol{\beta}(t) \neq \boldsymbol{c}(t), \tag{1.3}$$

where $\boldsymbol{C}$ is a given $q \times (p+1)$ full rank matrix, and $\boldsymbol{c}(t) = [c_1(t), \ldots, c_q(t)]^T$ a vector of given functions. This testing problem includes a lot of useful testing problems as special cases; the testing procedure proposed here works in general. For the GLHT problem (1.3), we adopt the F-type test proposed by Shen and Faraway (2004) for an important special case of (1.3) in which two nested FLMs are compared. The F-type test has two advantages: (1) it is scale-invariant and (2) its null distribution can be approximated by a usual F-statistic with degrees of freedom proportional to a constant κ that depends on the covariance function of the FLM only. This allows easy and fast implementation of the F-type test provided κ is properly estimated. The F-type test for the GLHT problem (1.3) warrants attention as Shen and Faraway (2004) did not study its asymptotic power. By studying asymptotic power, we show that the F-type test is root-n consistent. Shen and Faraway (2004) also adopted a naive method for estimating κ. That may not work well in some situations, and can be improved.

The major contributions of the paper include: (1) we establish the asymptotic power of the F-type test and show that under some mild conditions, the F-type test is root-n consistent; and (2) we propose a bias-reduced method to estimate κ, which helps improve the performance of the F-type test.

Significance tests for functional data have gotten attention recently. Among others, Faraway (1997) discussed the difficulties of extending multivariate hypothesis testing procedures to the context of functional data analysis and pointed out that the likelihood-based testing procedures might be less powerful. He proposed an L^2-norm based test and approximated the null distribution of the test statistic with the bootstrap. Ramsay and Silverman (1997) suggested a pointwise t-test or F-test but they did not discuss global tests. For curve data from stationary Gaussian processes, Fan and Lin (1998) developed some adaptive Neyman tests. Some other testing procedures, such as the Cramer-von Mises type test for two-sample problems for functional data, can be found in Hall and van Keilegom (2007) and reference therein.

In practice, functional data are observed discretely and with noise. When the noise variances over time are relatively small, and the functional responses of all the subjects are observed on a grid of evenly spaced design time points, say, $t_1, \ldots, t_M$ with a reasonably large M, the F-type test may be directly applied to the raw data as in Shen and Faraway (2004), Shen and Xu (2007) and Yang et al. (2007), among others. In other cases, individual functions can be reconstructed based on the observed discrete functional data set via such smoothing techniques as regression splines (Eubank (1999)), smoothing splines (Wahba (1990), Green and Silverman (1994)), P-splines (Ruppert, Wand, and Carroll (2003)), local polynomial smoothing (Wand and Jones (1995), Fan and Gijbels (1996)), and reproducing kernel Hilbert space decomposition (Wahba (1990), Ramsay and Dalzell (1991)) among others. Smoothing can remove most of the noise to allow the evaluation of individual functions at any resolution, and to generally improve the power of the F-type test. Zhang and Chen (2007) demonstrated how to reconstruct individual functions from a discrete functional data set using local polynomial smoothing. They showed that, under some mild conditions, the effects of substitutions of the functions with their local polynomial reconstructions can be ignored asymptotically. In this paper, we assume model (1.2) is true to ease presentation, while in practice, statistical inferences for functional linear models are based on the raw data or the reconstructed individual functions from the functional data set.

The rest of the paper is organized as follows. In Section 2, we describe the F-type test and derive its asymptotic power under a sequence of local alternatives. The naive method and the bias-reduced method for approximating the null distribution of the test statistic are also discussed in this section. Applications to the ergonomics data are presented in Section 3.1, and a simulation study is given in Section 3.2. Technical proofs of the main theoretical results are outlined in the Appendix.

2. The F-type Test

2.1. The test statistic

Let $\boldsymbol{X}=[\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n]^T, \boldsymbol{y}(t)=[y_1(t),\ldots,y_n(t)]^T$, and $\boldsymbol{v}(t)=[v_1(t),\ldots,v_n(t)]^T$ denote the design matrix, the response vector, and the subject-effect vector, respectively. Then the GLHT problem (1.3) can be written in a full-reduced model comparison format as

$$\begin{aligned} &H_0 : \boldsymbol{y}(t) = \boldsymbol{X}\boldsymbol{\beta}(t) + \boldsymbol{v}(t), \text{ subject to } \boldsymbol{C}\boldsymbol{\beta}(t) = \boldsymbol{c}(t), \\ &H_1 : \boldsymbol{y}(t) = \boldsymbol{X}\boldsymbol{\beta}(t) + \boldsymbol{v}(t). \end{aligned} \tag{2.1}$$

As in classical linear models, the models defined by H_0 and H_1 are referred to as the reduced model (RM) and the full model (FM) respectively. The RM is obtained from the FM by an additional linear constraint; here and throughout, we indicate a linear constraint with the subscript "c". Following Shen and Faraway (2004), who considered a special case of (2.1), we define the F-type test statistic

$$F_n = \frac{(\mathrm{ISE}_c - \mathrm{ISE})/q}{\mathrm{ISE}/(n-p-1)}, \tag{2.2}$$

where ISE_c and ISE denote the integrated squared errors under the RM and FM, respectively. Here, q and $n-p-1$ are related to, but no longer the degrees of freedom of $\mathrm{ISE}_c - \mathrm{ISE}$ and ISE, respectively.

Throughout, we assume that $\boldsymbol{X}$ has full rank. Then the least squares estimator of $\boldsymbol{\beta}(t)$ under the FM is $\hat{\boldsymbol{\beta}}(t) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}(t)$, which minimizes the integrated squared error

$$Q(\boldsymbol{\beta}) = \int_T \|\boldsymbol{y}(t) - \boldsymbol{X}\boldsymbol{\beta}(t)\|^2 dt, \tag{2.3}$$

where $\|\cdot\|$ denotes the usual L^2-norm of a vector. We take

$$\text{ISE} = \int_T \|\boldsymbol{y}(t) - \boldsymbol{X}\hat{\boldsymbol{\beta}}(t)\|^2 dt = \int_T \|\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t)\|^2 dt, \tag{2.4}$$

where $\hat{\boldsymbol{y}}(t) = \boldsymbol{X}\hat{\boldsymbol{\beta}}(t)$ denotes the fitted response vector under the FM. It follows that $\text{ISE} = \int_T \boldsymbol{y}(t)^T(\boldsymbol{I}_n - \boldsymbol{P}_X)\boldsymbol{y}(t)dt$, where $\boldsymbol{I}_n$ denotes the identity matrix of size n and $\boldsymbol{P}_X = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ denotes the projection matrix of the functional linear model regression. Similarly, we take $\text{ISE}_c = \int_T \|\boldsymbol{y}(t) - \boldsymbol{X}\hat{\boldsymbol{\beta}}_c(t)\|^2 dt = \int_T \|\boldsymbol{y}(t) - \hat{\boldsymbol{y}}_c(t)\|^2 dt$, where $\hat{\boldsymbol{y}}_c(t) = \boldsymbol{X}\hat{\boldsymbol{\beta}}_c(t)$ denotes the fitted response vector under the RM, with $\hat{\boldsymbol{\beta}}_c(t)$ being the least squares estimator of $\boldsymbol{\beta}(t)$ under the RM. By applying the Lagrangian multiplier method pointwisely, $\hat{\boldsymbol{\beta}}_c(t)$ can be obtained by minimizing the integrated squared errors (2.3) subject to the linear constraint $\boldsymbol{C}\boldsymbol{\beta}(t) = \boldsymbol{c}(t)$. Hence, it can be expressed as $\hat{\boldsymbol{\beta}}_c(t) = \hat{\boldsymbol{\beta}}(t) + \boldsymbol{h}(t)$, where $\boldsymbol{h}(t) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{C}^T]^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}}(t) - \boldsymbol{c}(t))$. Then we have $\hat{\boldsymbol{y}}_c(t) = \hat{\boldsymbol{y}}(t) + \boldsymbol{X}\boldsymbol{h}(t)$ or, equivalently, $\hat{\boldsymbol{v}}_c(t) = \hat{\boldsymbol{v}}(t) - \boldsymbol{X}\boldsymbol{h}(t)$, where $\hat{\boldsymbol{v}}_c(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{y}}_c(t)$ and $\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t) = (\boldsymbol{I}_n - \boldsymbol{P}_X)\boldsymbol{y}(t)$ denote the estimators of the subject-effect vector $\boldsymbol{v}(t)$ under the RM and FM, respectively. Since $(\boldsymbol{I}_n - \boldsymbol{P}_X)\boldsymbol{X} = 0$, $\hat{\boldsymbol{v}}(t)$ and $\boldsymbol{X}\boldsymbol{h}(t)$ are independent with their cross-product being 0. It follows that $\|\boldsymbol{y}(t) - \hat{\boldsymbol{y}}_c(t)\|^2 = \|\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t)\|^2 + \|\boldsymbol{X}\boldsymbol{h}(t)\|^2$. Therefore, we have the decomposition of the integrated squares $\text{ISE}_c = \text{ISE} + \text{ISH}$, where ISH denotes the extra integrated squares explained by the FM against the RM, and ISH and ISE are independent. It is seen that

$$\begin{aligned} \text{ISH} = \text{ISE}_c - \text{ISE} &= \int_T \|\boldsymbol{X}\boldsymbol{h}(t)\|^2 dt \\ &= \int_T \left[\boldsymbol{C}\hat{\boldsymbol{\beta}}(t) - \boldsymbol{c}(t)\right]^T \left[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^T\right]^{-1} \left[\boldsymbol{C}\hat{\boldsymbol{\beta}}(t) - \boldsymbol{c}(t)\right] dt \\ &= \int_T \|\boldsymbol{w}(t)\|^2 dt, \end{aligned} \tag{2.5}$$

where $\boldsymbol{w}(t) = \left[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^T\right]^{-1/2}\left[\boldsymbol{C}\hat{\boldsymbol{\beta}}(t) - \boldsymbol{c}(t)\right]$. Notice that the above ISH is the test statistic of the L^2-norm based test of Zhang and Chen (2007), and

$$\boldsymbol{w}(t) \sim \text{GP}(\boldsymbol{\eta}_w, \boldsymbol{\gamma}_w), \tag{2.6}$$

where $\boldsymbol{\eta}_w(t) = \left[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^T\right]^{-1/2}\left[\boldsymbol{C}\boldsymbol{\beta}(t) - \boldsymbol{c}(t)\right]$ and $\boldsymbol{\gamma}_w(s,t) = \gamma(s,t)\boldsymbol{I}_q$.

We introduce some notation. By $W \stackrel{d}{=} Z$, we mean that W and Z have the same distribution; when $W \sim \chi_d^2(w^2)$, we mean that W is a chi-squared random variable with d degrees of freedom and a noncentrality parameter w^2; $\mathrm{tr}(\gamma) = \int_T \gamma(t,t)dt$ denotes the trace of $\gamma(s,t)$. We collect two regularity assumptions.

Assumption A. (1) $0 < \mathrm{tr}(\gamma) < \infty$, and (2) as $n \to \infty$, $n^{-1}\boldsymbol{X}^T\boldsymbol{X} \to \boldsymbol{\Omega}$ with $\boldsymbol{\Omega}$ invertible.

Under A(1) $\gamma(s,t)$ has finite trace, hence (see Wahba (1990))

$$\gamma(s,t) = \sum_{r=1}^{m} \lambda_r \phi_r(s)\phi_r(t), \tag{2.7}$$

where $\lambda_1, \lambda_2, \cdots, \lambda_m$ are the decreasingly-ordered positive eigenvalues of $\gamma(s,t)$, and $\phi_1(t), \phi_2(t), \ldots, \phi_m(t)$ are the associated orthonormal eigenfunctions. Set $m = \infty$ when all the eigenvalues are positive. By (2.7), it is easy to show that

$$\mathrm{tr}(\gamma) = \sum_{r=1}^{m} \lambda_r, \quad \mathrm{tr}(\gamma^{\otimes 2}) = \sum_{r=1}^{m} \lambda_r^2 \le \mathrm{tr}^2(\gamma) < \infty, \tag{2.8}$$

where $\gamma^{\otimes 2}(s,t) = \int_T \gamma(s,u)\gamma(u,t)du$. The inequality in (2.8) is obvious when one notices that $\sum_{r=1}^m \lambda_r^2 \le (\sum_{r=1}^m \lambda_r)^2$.

Theorem 1. *Under Assumption A, we have*

$$ISH \stackrel{d}{=} \sum_{r=1}^{m} \lambda_r A_r + \sum_{r=m+1}^{\infty} \pi_r^2, \quad \mathrm{ISE} \stackrel{d}{=} \sum_{r=1}^{m} \lambda_r B_r, \tag{2.9}$$

where A_r, B_r *are independent with* $A_r \sim \chi_q^2(\lambda_r^{-1}\pi_r^2)$, $B_r \sim \chi_{n-p-1}^2$, *and* $\pi_r^2 = \| \int_T \boldsymbol{\eta}_w(t)\phi_r(t)dt\|^2$, $r = 1, \ldots, m$. *It follows that*

$$F_n \stackrel{d}{=} \frac{\left(\sum_{r=1}^m \lambda_r A_r + \sum_{r=m+1}^\infty \pi_r^2\right)/q}{\sum_{r=1}^m \lambda_r B_r/(n-p-1)}. \tag{2.10}$$

Notice that as $n \to \infty$, the denominator of F_n tends to $\mathrm{tr}(\gamma)$ almost surely. Let F_n^* denote F_n under H_0. Then by Theorem 1 it is easy to see that, as $n \to \infty$, we have

$$F_n^* \stackrel{d}{=} \frac{\sum_{r=1}^m \lambda_r A_r/q}{\sum_{r=1}^m \lambda_r B_r/(n-p-1)} \stackrel{d}{=} \frac{T^*}{q\,\mathrm{tr}(\gamma)} + o_p(1), \tag{2.11}$$

where $T^* = \sum_{r=1}^m \lambda_r A_r$, $A_r \stackrel{i.i.d.}{\sim} \chi_q^2$. Thus, under Assumption A(1), T^* has finite mean $q\,\mathrm{tr}(\gamma)$ and finite variance $2q\,\mathrm{tr}(\gamma^{\otimes 2})$.

2.2. Asymptotic power under local alternatives

In this subsection, we investigate the asymptotic power of F_n. When an alternative is fixed, it is easy to show that the associated power tends to 1 as $n \to \infty$. We study the power of F_n when alternatives tend to the null hypothesis at a rate slightly slower than root-n. For this purpose, a sequence of local alternatives is

$$H_{1n} : \boldsymbol{C\beta}(t) - \boldsymbol{c}(t) = n^{-\tau/2}\boldsymbol{d}(t), \tag{2.12}$$

where τ is some constant and $\boldsymbol{d}(t)$ is any fixed real vector of functions satisfying the following.

Assumption B. $0 \le \tau < 1$, and $0 < \int_T \|\boldsymbol{d}(t)\|^2 dt < \infty$.

Under Assumption A, H_{1n}, and by Theorem 1, we have $\pi_r^2 = n^{1-\tau}\delta_r^2$ where $\delta_r^2 = \| \int_T \left(\boldsymbol{C\Omega}^{-1}\boldsymbol{C}^T\right)^{-1/2} \boldsymbol{d}(t)\phi_r(t)dt\|^2[1+o(1)]$, $r = 1, \ldots$. Notice that there are only two cases for us to consider: (1) $\delta_r^2 = 0$ for all $r \in \{1, \ldots, m\}$, and (2) $\delta_r^2 \ne 0$ for at least one $r \in \{1, \ldots, m\}$. Case (1) implies that m is finite and describes a case of no information about the violation of H_0 projected onto the subspace spanned by all the m eigenfunctions with positive eigenvalues. We show that the asymptotic power of F_n under H_{1n} tends to 1 as $n \to \infty$ in both cases. Therefore, the F-type test is root n-consistent. That is, the F-type test can effectively detect departures from the null hypothesis of size $n^{-1/2}$ in the direction of any given vector of functions satisfying Assumption B. This shows that the F-type test is reasonably powerful for general functional linear hypothesis testing problems.

We consider the asymptotic power of F_n under Case (1). Here we can see that $F_n \overset{d}{=} \{\left[\sum_{r=1}^m \lambda_r A_r + n^{1-\tau}\delta^2\right]/q\}/\{\sum_{r=1}^m \lambda_r B_r/(n-p-1)\} \overset{d}{=} \{T^* + n^{1-\tau}\delta^2\}/\{q\,\mathrm{tr}(\gamma)\} + o_p(1)$, where T^* is as defined in (2.11) and

$$\delta^2 = \sum_{r=1}^{\infty} \delta_r^2 = \int_T \boldsymbol{d}(t)^T \left(\boldsymbol{C\Omega}^{-1}\boldsymbol{C}^T\right)^{-1} \boldsymbol{d}(t)dt > 0, \tag{2.13}$$

since $\boldsymbol{C\Omega}^{-1}\boldsymbol{C}^T$ is positive definite. For any $\alpha \in (0,1)$, let $F_{n,\alpha}^*$ and T_α^* be the upper 100α percentiles of F_n^* and T^*, respectively. Then by (2.11), we have $F_{n,\alpha}^* = \{T_\alpha^*\}/\{q\,\mathrm{tr}(\gamma)\} + o_p(1)$.

Theorem 2. *For Case* (1) *under Assumptions A and B, the asymptotic power of F_n is*

$$P\left(F_n \ge F_{n,\alpha}^* | H_{1n}\right) = P\left(T^* > T_\alpha^* - n^{1-\tau}\delta^2\right) + o(1),$$

which tends to 1 *as $n \to \infty$.*

We now investigate the asymptotic power of F_n under Case (2). Here we have $\delta_\lambda^2 = \sum_{r=1}^m \lambda_r \delta_r^2 > 0$. Let $\Phi(\cdot)$ denote the cumulative distribution function of the standard normal distribution $N(0,1)$.

Theorem 3. *For Case* (2) *under Assumptions A and B, as $n \to \infty$, we have*

$$\left\{F_n - \frac{n^{1-\tau}\delta^2}{q\,\mathrm{tr}(\gamma)}\right\} \Big/ \left\{\frac{2n^{(1-\tau)/2}\delta_\lambda}{q\,\mathrm{tr}(\gamma)}\right\} \xrightarrow{L} N(0,1). \tag{2.14}$$

In addition, the asymptotic power of F_n is

$$P\left(F_n \geq F^*_{n,\alpha}|H_{1n}\right) = \Phi\left(\frac{n^{(1-\tau)/2}\delta^2}{2\delta_\lambda}\right) + o(1), \tag{2.15}$$

which tends to 1 *as $n \to \infty$.*

Theorem 2 shows that under Case (1), the F-type test is still powerful in detecting the violation of the null hypothesis; dimension-reduction based testing procedures may be less powerful. Moreover, it, together with Theorem 3, shows that the asymptotic power of F_n increases to 1 at a rate $O(n^{1-\tau})$ under Case (1), much faster than the rate $O(n^{(1-\tau)/2})$ under Case (2), as $n \to \infty$. This is reasonable since, from Theorem 3, it is seen that the asymptotic variance of F_n under Case (2) is also increasing at a rate $O(n^{(1-\tau)/2})$ as n increases, while this is not the case under Case (1). Notice that it is easy to show that Theorems 2 and 3 still hold when the critical value $F^*_{n,\alpha}$ is replaced with some estimated critical value.

2.3. Null distribution approximation

In this subsection, we point out that the null distribution of the F-type test statistic F_n in (2.2) can be well-approximated by that of an F-distribution with $q\kappa$ and $(n-p-1)\kappa$ degrees of freedom, where κ is a constant purely determined by the underlying covariance function $\gamma(s,t)$. That is to say, under H_0 we have

$$F_n = \frac{(\mathrm{ISE}_c - \mathrm{ISE})/q}{\mathrm{ISE}/(n-p-1)} = \frac{\mathrm{ISH}/(q\kappa)}{\mathrm{ISE}/[(n-p-1)\kappa]} \overset{approx.}{\sim} F_{q\kappa,(n-p-1)\kappa}. \tag{2.16}$$

Therefore the F-type test can be easily conducted if κ is properly estimated.

First notice that by Theorem 1, ISH and ISE are independent and, under H_0, they are central χ^2-type mixtures whose distributions can be well-approximated by the χ^2-approximation that matches two or three cumulants (Shen and Faraway (2004), Zhang (2005)). Here, due to its simplicity, the two-cumulant matched χ^2-approximation method of Shen and Faraway (2004) is adopted to approximate the null distributions of ISH and ISE, respectively, so that we get (2.16) straightforwardly with

$$\kappa = \frac{\left(\sum_{r=1}^m \lambda_r\right)^2}{\sum_{r=1}^m \lambda_r^2} = \frac{\mathrm{tr}^2(\gamma)}{\mathrm{tr}(\gamma^{\otimes 2})}, \tag{2.17}$$

where λ_r, $r = 1, \ldots, m$ are the nonzero eigenvalues of $\gamma(s,t)$ as defined before.

It is clear that the constant κ is important. Shen and Faraway (2004) called κ as the "degrees of freedom adjustment factor". Notice that κ may not always be an integer, hence the same for $q\kappa$ and $(n-p-1)\kappa$. This is not a problem since popular statistical software such as Matlab allow non-integer degrees of freedom for the F-distribution. It may be a problem for those users who conduct the F-type test by looking at the F-table for proper critical values, but Shen and Faraway (2004) proposed truncating the approximate degrees of freedoms d_1 and d_2 to arrive at $F_{[d_1],[d_2]}$ where $[a]$ denotes the closest integer to a. When both d_1 and d_2 are large, this method works well, but may be misleading otherwise.

From (2.17), we can see that to estimate κ, we do not need to estimate m and the m nonzero eigenvalues $\lambda_1, \cdots, \lambda_m$ of $\gamma(s,t)$, a difficult task, especially when $\gamma(s,t)$ is evaluated with very high resolution. Rather, we need first to estimate $\gamma(s,t)$. By Lemma 2 and the proof of Theorem 1 in the Appendix, $\gamma(s,t)$ can be unbiasedly estimated by

$$\hat{\gamma}(s,t) = \sum_{i=1}^{n} (y_i(s) - \hat{y}_i(s)) \frac{(y_i(t) - \hat{y}_i(t))}{n-p-1}. \tag{2.18}$$

Shen and Faraway (2004) proposed estimating κ by replacing the eigenvalues λ_r, $r = 1, \ldots$ of $\gamma(s,t)$ in the expression of κ with the eigenvalues $\hat{\lambda}_r$, $r = 1, \ldots$ of $\hat{\gamma}(s,t)$. The resulting estimator is $\hat{\kappa}_0 = \mathrm{tr}^2(\hat{\gamma})/\mathrm{tr}(\hat{\gamma}^{\otimes 2})$. By some simple algebra, one can show that

$$1 \le \hat{\kappa}_0 \le \hat{m} \le n-p-1, \tag{2.19}$$

where $\hat{m}$ is the number of nonzero eigenvalues of $\hat{\gamma}(s,t)$. When $\hat{m} = 1$, the first equality holds, and when all the nonzero eigenvalues are equal the second equality holds. Note that $\hat{\kappa}_0$ is biased since we can show that $\mathrm{tr}^2(\hat{\gamma})$ and $\mathrm{tr}(\hat{\gamma}^{\otimes 2})$ are biased for $\mathrm{tr}^2(\gamma)$ and $\mathrm{tr}(\gamma^{\otimes 2})$ respectively; see Lemma 1 in the Appendix for more details. We propose a bias-reduced method that replaces $\mathrm{tr}^2(\gamma)$ and $\mathrm{tr}(\gamma^{\otimes 2})$ with their unbiased estimators.

The unbiased estimators of $\mathrm{tr}^2(\gamma)$ and $\mathrm{tr}(\gamma^{\otimes 2})$ can be obtained by applying Lemma 2 given in the Appendix. Using $\hat{\gamma}(s,t)$ defined in (2.18), the unbiased estimators are, respectively,

$$\begin{aligned} \widehat{\mathrm{tr}^2(\gamma)} &= \frac{(n-p-1)(n-p)}{(n-p-2)(n-p+1)} \left[\mathrm{tr}^2(\hat{\gamma}) - \frac{2}{n-p} \mathrm{tr}(\hat{\gamma}^{\otimes 2}) \right], \\ \widehat{\mathrm{tr}(\gamma^{\otimes 2})} &= \frac{(n-p-1)^2}{(n-p-2)(n-p+1)} \left[\mathrm{tr}(\hat{\gamma}^{\otimes 2}) - \frac{1}{n-p-1} \mathrm{tr}^2(\hat{\gamma}) \right]. \end{aligned} \tag{2.20}$$

It follows that the bias-reduced estimator of κ is

$$\hat{\kappa} = \frac{(n-p)\left[\mathrm{tr}^2(\hat{\gamma}) - 2\,\mathrm{tr}(\hat{\gamma}^{\otimes 2})/(n-p)\right]}{(n-p-1)\left[\mathrm{tr}(\hat{\gamma}^{\otimes 2}) - \mathrm{tr}^2(\hat{\gamma})/(n-p-1)\right]} = \frac{(n-p)\hat{\kappa}_0 - 2}{(n-p-1) - \hat{\kappa}_0}. \tag{2.21}$$

Since $\gamma(s,t)$ has at least one nonzero eigenvalue, $\mathrm{tr}(\gamma^{\otimes 2}) > 0$. It follows from the second expression in (2.20) that $\hat{\kappa}_0 < n-p-1$ almost surely. In fact, in data analysis we always have $\hat{\kappa}_0 < n-p-1$. Were $\hat{\kappa}_0 = n-p-1$, we would have $\hat{m} = n-p-1$, and all the $n-p-1$ nonzero eigenvalues would be equal; in data analysis, the nonzero eigenvalues of $\hat{\gamma}(s,t)$ are not the same even if the nonzero eigenvalues of $\gamma(s,t)$ are equal. Thus, without loss of generality, we assume that $\hat{\kappa}_0 < n-p-1$. This, together with (2.19), leads to

$$\hat{\kappa} - \hat{\kappa}_0 = \frac{(\hat{\kappa}_0 + 2)(\hat{\kappa}_0 - 1)}{(n-p-1) - \hat{\kappa}_0} \geq 0,$$

with the equality holding when $\hat{m} = 1$. Then the difference between $\hat{\kappa}$ and $\hat{\kappa}_0$ decreases with increasing n.

The bias-reduced method can be regarded as a generalization of the bias-reduced method used by Huynh and Feldt (1976) in the randomized block and split-plot designs for repeated measurements. Both methods aim to reduce the biases of the degrees of freedom adjustment factors for their associated F-type tests.

While $\hat{\kappa}$ is still biased, its bias is largely reduced. A simulation study presented in Section 3.2 indicates that the bias-reduced method and the naive method perform similarly when the data are highly or moderately correlated and the former is preferred when the data are nearly uncorrelated. This simulation result is similar to the one obtained by Huynh and Feldt (1976).

3. Numerical Results

3.1. Applications to the ergonomics data

For convenience, we refer to the F-type test with the bias-reduced method as the bias-reduced F-type test. In this subsection, we illustrate the bias-reduced F-type test using the ergonomics data introduced in Section 1. Following Shen and Faraway (2004), we first removed Curve 37 since the subject changed his mind about the target location in mind-reach so that the associated motion to the left rear shifter location was clearly wrong. We then fit the angle curves using a quadratic regression spline with 7 equally spaced inner knots. The number of knots was selected by GCV (Zhang and Chen (2007)). For practical computation, we evaluated the individual curves at a grid of M equally spaced time points over the support of the design time points, e.g., $[0,1]$, for the ergonomics data. The number M must be large enough to give good approximations to the integrals involved in the ISE and ISE_c of Section 2. For all the numerical results presented in this section, we took $M = 1{,}000$; use of larger M increased the computation time substantially, but generally did not give more precise results. By the bias-reduced method, we found that $\hat{\kappa} = 1.81$ based on the full quadratic model (1.1) of Section 1 for the ergonomics data. This estimate of κ is used in all the tests below.

We first conducted an overall test to check if the quadratic model (1.1) was statistically significant. We took $\boldsymbol{C} = [\boldsymbol{0}, \boldsymbol{I}_p]$ and $\boldsymbol{c}(t) = \boldsymbol{0}$ to calculate the F-type test statistic F_n defined in (2.2), to get $\hat{F}_n = 31.40$. Since $n = 60 - 1 = 59, p = 10$, and $q = p - 1 = 9$, the approximated degrees of freedom of the numerator and denominator of the test statistic F_n were $\mathrm{df}_1 = q\hat{\kappa} = 16.29$ and $\mathrm{df}_2 = (n - p - 1)\hat{\kappa} = 86.88$. The associated P-value was 0, indicating that the overall test was highly significant.

We then tested whether each of the coefficient functions was statistically significant. For this end, we took $\boldsymbol{C} = \boldsymbol{e}_{j,p+1}^T$ and $\boldsymbol{c}(t) = 0$ for $j = 0, \ldots, p$, respectively. The calculated F_n are listed in the second column of Table 1. Since $q = 1$, we found $\mathrm{df}_1 = \hat{\kappa} = 1.81$ and $\mathrm{df}_2 = (n - p - 1)\hat{\kappa} = 86.88$. The associated P-values are listed in the third column of Table 1. It is seen that all the coefficient functions are significant or highly significant except the coefficient function $\beta_8(t)$. Figure 1 displays the four estimated coefficient functions (solid curves) with their 95% pointwise confidence bands (dashed curves). From the left lower panel, it is seen that the pointwise confidence band for $\beta_8(t)$ indeed contains "0" most of time, indicating that one may delete $\beta_8(t)$ from the model so that a more parsimonious model can be obtained.

Should we remove the insignificant coefficient function $\beta_8(t)$ from the full quadratic model (1.1)? Or generally speaking, should we delete those insignificant coefficient functions in a functional linear model? Unfortunately, there is no simple answer. Some believe that, for a functional linear model, it makes more sense to keep all or none of a set of terms of the same order, while others believe that one should delete those insignificant coefficient functions and just retain those significant coefficient functions to yield a parsimonious and efficient model in order, to improve the prediction performance of the predictors, and to obtain a better understanding of the underlying process that generates the data. Although this problem warrants further study, we present an illustration of how to do variable selection in functional linear models, as follows.

Various variable selection methods, e.g., forward, backward, and stepwise procedures in classical linear regression models can be naturally adopted for functional linear models with the usual F-test replaced by our bias-reduced F-type test. Here, as an illustration, we adopted the backward selection method for the quadratic model (1.1). That is, we deleted one least significant coefficient function step by step until all the coefficient functions were significant. The significance of a variable is usually determined by the significant level specified for variable selection, and this is usually larger than the significant level, e.g., 5%, specified for hypothesis testing. Here for simplicity, we took the significant

Table 1. Coefficient significance table for the quadratic model (1.1) for the ergonomics data.

Estimated Coef. function	F	P-value
$\hat{\beta}_0(t)$	2177.8	0
$\hat{\beta}_1(t)$	12.64	2.86×10^{-5}
$\hat{\beta}_2(t)$	24.86	8.92×10^{-9}
$\hat{\beta}_3(t)$	7.44	1.48×10^{-3}
$\hat{\beta}_4(t)$	7.30	1.65×10^{-3}
$\hat{\beta}_5(t)$	10.63	1.26×10^{-4}
$\hat{\beta}_6(t)$	3.37	4.32×10^{-2}
$\hat{\beta}_7(t)$	8.08	8.88×10^{-4}
$\hat{\beta}_8(t)$	1.45	$\mathbf{2.40\times10^{-1}}$
$\hat{\beta}_9(t)$	4.88	1.20×10^{-2}

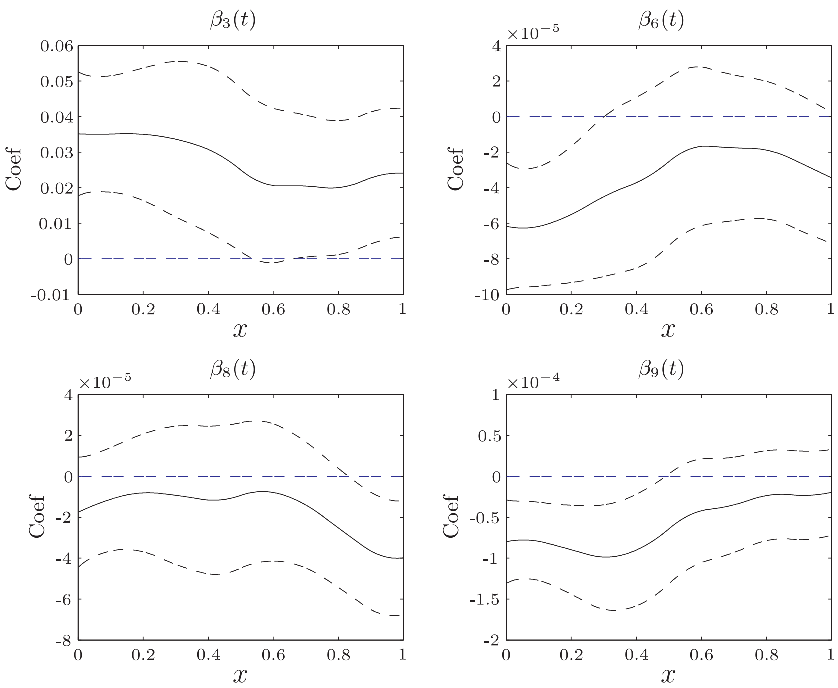

Figure 1. Four estimated coefficient functions (solid curves) with 95% pointwise confidence bands (dashed curves) under the full quadratic model (1.1).

level for variable selection as 5% for illustrative purpose only. Table 2 shows that the P-values for each step. The second column shows the P-values for the coefficient functions in Step 1, that led us to delete the term a_ic_i (i.e., $\beta_8(t)$) from the quadratic model (1.1). The P-values in the third and fourth columns for Steps 2 and 3 led us to delete the terms c_i^2 (i.e., $\beta_6(t)$) and b_ic_i (i.e., $\beta_9(t)$),

Table 2. P-values for variable selection based on the quadratic model (1.1) for the ergonomics data.

Estimated Coef. function	P-values for Step 1	2	3	4
$\hat{\beta}_0(t)$	0	0	0	0
$\hat{\beta}_1(t)$	2.86×10^{-5}	4.58×10^{-5}	4.90×10^{-5}	7.44×10^{-5}
$\hat{\beta}_2(t)$	8.92×10^{-9}	1.64×10^{-8}	1.48×10^{-11}	0
$\hat{\beta}_3(t)$	1.48×10^{-3}	2.60×10^{-3}	2.71×10^{-3}	1.37×10^{-5}
$\hat{\beta}_4(t)$	1.65×10^{-3}	3.99×10^{-3}	1.26×10^{-2}	1.88×10^{-2}
$\hat{\beta}_5(t)$	1.26×10^{-4}	3.13×10^{-4}	9.30×10^{-6}	3.63×10^{-8}
$\hat{\beta}_6(t)$	4.32×10^{-2}	$\mathbf{6.85\times10^{-2}}$		
$\hat{\beta}_7(t)$	8.88×10^{-4}	2.40×10^{-3}	2.94×10^{-3}	4.66×10^{-3}
$\hat{\beta}_8(t)$	$\mathbf{2.40\times10^{-1}}$			
$\hat{\beta}_9(t)$	1.20×10^{-2}	1.72×10^{-2}	$\mathbf{1.36\times10^{-1}}$	

respectively. As shown by the P-values in the fifth column, we did not delete any further coefficient function. We then obtained the final reduced model

$$\begin{aligned} y_i(t) = \beta_0(t) + a_i\beta_1(t) + b_i\beta_2(t) + c_i\beta_3(t) + a_i^2\beta_4(t) + b_i^2\beta_5(t) \\ +a_ib_i\beta_7(t) + v_i(t),\ i = 1,\ldots,60. \end{aligned} \tag{3.1}$$

From (3.1), we can see that the angle curve $y(t)$ has a significant quadratic relationship with the "left to right" direction a and the "close to far" direction b, but only has a significant linear relationship with the "down to up" direction c.

To test if the final reduced model (3.1) is adequate against the quadratic model (1.1), we conducted a bias-reduced F-type test taking

$$\boldsymbol{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \text{ and } \boldsymbol{c}(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

to test if $\beta_6(t), \beta_8(t)$ and $\beta_9(t)$ were significant. The resulting P-value was 0.0609, indicating that the final reduced model (3.1) is nearly adequate.

3.2. A Simulation study

For simplicity, we denote the L^2-norm based test of Zhang and Chen (2007), the F-type test of Shen and Faraway (2004), and the bias-reduced F-type test by L_0^2, F_0, and F_1, respectively. In this subsection, we present a simulation study to investigate the finite-sample performance of these testing procedures. We first compare F_1 and F_0 and then F_0 and L_0^2.

We generated samples from the FLM (1.2) described in Section 1. For simplicity we adopted the design time points $\boldsymbol{x}_i, i = 1, \ldots, n$, of the ergonomics data and took the estimated $\hat{\boldsymbol{\beta}}(t)$ as the underlying $\boldsymbol{\beta}(t)$ except for letting

$$[\beta_6(t), \beta_8(t), \beta_9(t)]^T = \Delta \times \left[\hat{\beta}_6(t), \hat{\beta}_8(t), \hat{\beta}_9(t)\right]^T, \quad \text{where } \Delta \in [0, \Delta_0].$$

This allows us to compare the powers of the three testing procedures for testing $H_0 : [\beta_6(t), \beta_8(t), \beta_9(t)]^T = \mathbf{0}$. Notice that when $\Delta = 0$, H_0 holds and when Δ increases, the powers of the testing procedures also increase. For easy presentation, we chose Δ_0 so that the powers of the testing procedures at Δ_0 were about 1. Let m_0 be an odd positive integer. To control the correlation of the subject-effect functions, we generated $v_i(t)$ using $v_i(t) = \sum_{r=1}^{m_0} \xi_{ir}\psi_r(t),\ i = 1, \ldots, n$, where the basis functions $\psi_r(t), r = 1, \ldots, m_0$ were orthonormal and smooth over the support of the ergonomics data, and the coefficients $\xi_{ir} \sim N(0, \lambda_r), r = 1, \ldots, m_0$, independent of each other, so that the associated covariance function of $v_i(t)$ was $\gamma(s, t) = \sum_{r=1}^{m_0} \lambda_r\psi_r(s)\psi_r(t)$. For simplicity, we used $\psi_1(t) = 1, \psi_{2r}(t) = \sqrt{2}\sin(2\pi rt), \psi_{2r+1}(t) = \sqrt{2}\cos(2\pi rt)$, $t \in [0, 1]$, $r = 1, \ldots, (m_0 - 1)/2$. We also specified $\lambda_r = a\rho^r$, $r = 1, \ldots, m_0$, for some $a > 0$ and $0 < \rho < 1$. Notice that ρ not only determines the decay rate of λ_r, $r = 1, \ldots, m_0$, but also determines how the resulting subject-effect functions $v_i(t),\ i = 1, \ldots, n$ are correlated: when ρ is close to 0 (resp. 1), the subject-effect functions are highly correlated (resp. nearly uncorrelated). We took $a = 1, m_0 = 21$, and $\rho = 0.100, 0.272, 0.444, 0.616, 0.788$ and 0.960, representing correlations from high to low. For each fixed ρ, we let Δ take a grid of equally spaced values in $[0, \Delta_0]$. For each pair (ρ, Δ), $N = 10{,}000$ samples were generated. For each sample, the test statistics of the testing procedures were computed and the associated P-values were calculated using the testing procedures under consideration. When the P-values were smaller than the nominal significance level α (5% here), the null hypothesis was rejected. The power of a testing procedure is the proportion of the number of rejections based on the calculated P-value using the associated testing procedure.

We first compare the bias-reduced method and the naive method. Figure 2 displays the power functions of the three testing procedures: L_0^2 (dotted), F_0 (dashed), and F_1 (solid). From Panels (a), (b), (c), and (d), it is seen that when the data were highly or moderately correlated ($\rho = 0.100, 0.272, 0.444$ and 0.616), the powers of F_1 and F_0 are about the same, and their Type-I errors (powers at $\Delta = 0$) are close to 5%, the nominal significance level, with the powers of F_1 slightly larger than those of F_0. Therefore, in these four cases, the difference between F_1 and F_0 is quite small. However, from Panels (e) and (f), it is seen that when the data were nearly uncorrelated ($\rho = 0.788$ and 0.960), the powers of F_1 were larger than those of F_0. In addition, the Type-I errors of F_1 were close

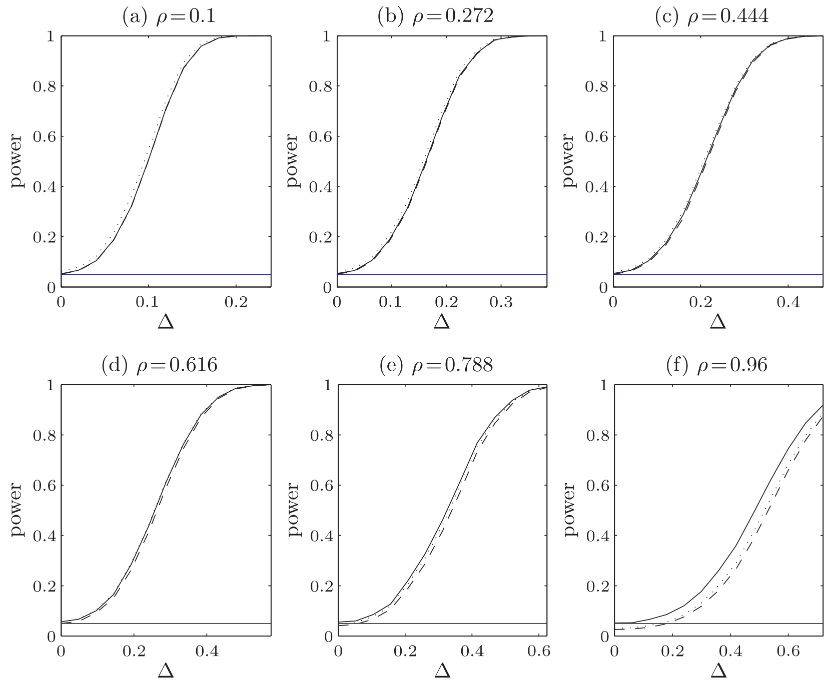

Figure 2. Power functions of L_0^2 (dotted), F_0 (dashed) and F_1 (solid) for various values of ρ.

to 5% while the Type-I errors of F_0 were lower than 5%. Thus, F_1 is preferred to F_0. That is, the bias-reduced method is preferred to the naive method here, especially when the data are nearly uncorrelated. This conclusion can be seen more clearly from Figure 3 below.

Figure 3 displays the absolute relative error (ARE) curves of the three testing procedures: L_0^2 (dotted), F_0 (dashed), and F_1 (solid). The ARE of a testing procedure is defined as $\{|\text{power} - \text{simulated power}|\}/\{\text{simulated power}\} \times 100$. The simulated power of a testing procedure is defined as the proportion of the number of rejections (out of N replications) based on the simulated critical values which are the upper 100α-percentiles of the N test statistics, computed based on the N samples generated with $\Delta = 0$ and the same ρ. The true underlying power of a testing procedure is not available but is estimated by the simulated power of the testing procedure when N is large.

From Panels (a), (b), (c), and (d), it is seen that although the AREs of F_0 are generally smaller than those of F_1 especially when the data are moderately correlated, the AREs of F_1 and F_0 are quite close to each other. Moreover, from Panels (e) and (f), it is seen that the AREs of F_1 are much smaller than those of F_0. It seems that the bias-reduced method is preferred to the naive method,

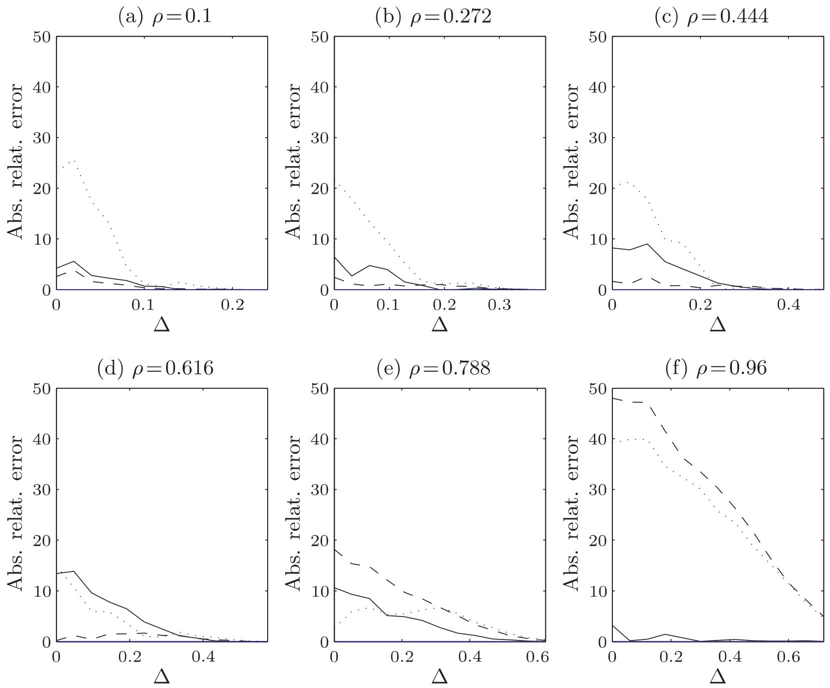

Figure 3. ARE curves of L_0^2 (dotted), F_0 (dashed) and F_1 (solid) for various values of ρ.

Table 3. Means (standard deviations) of ASRE($\hat{\kappa}$) over the values of Δ for the six cases.

ρ	0.100	0.272	0.444	0.616	0.788	0.960
Naive	0.0026(0.0000)	0.0090(0.0001)	0.0111(0.0002)	0.0125(0.0002)	0.0253(0.0002)	0.0880(0.0001)
Bias-reduced	0.0032(0.0001)	0.0116(0.0002)	0.0140(0.0003)	0.0117(0.0001)	0.0075(0.0001)	0.0021(0.0000)

especially when data are nearly uncorrelated. This result may be partially explained by the fact that the bias-reduced method generally resulted in better estimates for κ in the simulation study than did the naive method as shown in Table 3 below.

The accuracy of $\hat{\kappa}$ can be measured by its average squared relative error: $ASRE(\hat{\kappa}) = N^{-1}\sum_{i=1}^{N}(\hat{\kappa}_i/\kappa_i - 1)^2$, where N is the number of replications in the simulation study, and $\hat{\kappa}_i$ and κ_i are the estimates and true values of κ in the i-th replication. The true values of κ can be calculated using (2.17) with the true values of $\gamma(s,t)$. Table 3 lists the means (standard deviations) of ASRE($\hat{\kappa}$) over the values of Δ (values of ASRE($\hat{\kappa}$) are slightly different for different Δ) for the six cases under consideration. It is seen that, for the first four cases, the means of ASRE($\hat{\kappa}$) for the bias-reduced method and the naive method are of the same magnitude, although the means of ASRE($\hat{\kappa}$) for the naive method in the first three cases are slightly smaller than those for the bias-reduced method. It is also seen that, for the last two cases, the means of ASRE($\hat{\kappa}$) for the naive method are much larger than those for the bias-reduced method. In addition, the means of ASRE($\hat{\kappa}$) for the naive method increased with larger ρ, while the means of ASRE($\hat{\kappa}$) for the bias-reduced method were similar.

We now compare the F-type test and the L^2-norm based test. From Figure 2, it is seen that the powers of L_0^2 (dotted) and F_0 (dashed) are about the same, with the former slightly larger than the latter. From Figure 3, it is seen that the AREs of L_0^2 (dotted) are larger than those of F_0 (dashed) except in Panels (e) and (f) where the AREs of L_0^2 are slightly smaller than those of F_0. We might conclude that F_0 is slightly preferred to L_0^2, especially when the data are highly or moderately correlated. From Panel (f) of Figure 2, the Type-I errors of both F_0 and L_0^2 are lower than 5%, simply due to the fact that both F_0 and L_0^2 adopt the naive method.

Acknowledgements

The work was supported by the National University of Singapore Academic Research Grant R-155-000-085-112. The author thanks the co-Editor, an associate editor, and two reviewers for their constructive comments and invaluable suggestions that helped improve the paper substantially. He also thanks his research assistant Ms Jing Han for editing the tables used in this paper.

Appendix: Proofs

In this appendix, we first present two lemmas about Wishart processes and then give the technical proofs of Theorems 1 and 3.

A. Two lemmas about Wishart processes

Wishart processes are natural generalizations of Wishart random variables. Throughout, we use $\mathrm{WP}(n,\gamma)$ to denote a Wishart process with n degrees of freedom and a covariance function $\gamma(s,t)$. A Wishart process $W(s,t) \sim \mathrm{WP}(n,\gamma)$ can be written as

$$W(s,t) = \sum_{i=1}^{n} u_i(s)u_i(t), \tag{A.1}$$

where $u_i(t)$, $i = 1,\ldots,n$ are i.i.d. $\mathrm{GP}(0,\gamma)$.

Notice that when $\gamma(s,t)$ has finite trace, it has the singular value decomposition (2.7). It follows that the $u_i(t)$'s in (A.1) have the Karhunen-Loeve expansions

$$u_i(t) = \sum_{r=1}^{\infty} \xi_{ir}\phi_r(t), i = 1,\ldots,n, \tag{A.2}$$

where ξ_{ir} are independent, $\xi_{ir} \sim N(0, \lambda_r)$, and $\lambda_r, \phi_r(t)$ are the r-th eigenvalue and eigen-function of $\gamma(s,t)$ as defined in (2.7).

Lemma 1. *Assume* $W(s,t) \sim \mathrm{WP}(n,\gamma)$ *with* $\mathrm{tr}(\gamma) < \infty$. *Then we have*

(a) $\mathrm{E}W(s,t) = n\gamma(s,t)$,

(b) $\mathrm{tr}(W) \stackrel{d}{=} \sum_{r=1}^{\infty} \lambda_r A_r,\ A_r \stackrel{i.i.d.}{\sim} \chi_n^2$,

(c) $\mathrm{E}\,\mathrm{tr}(W) = n\,\mathrm{tr}(\gamma)$ *and* $\mathrm{Etr}^2(W) = 2n\,\mathrm{tr}(\gamma^{\otimes 2}) + n^2\mathrm{tr}^2(\gamma)$,

(d) $\mathrm{E}\,\mathrm{tr}(W^{\otimes 2}) = n(n+1)\,\mathrm{tr}(\gamma^{\otimes 2}) + n\mathrm{tr}^2(\gamma)$.

Proof of Lemma 1. Let $W_i(s,t) = u_i(s)u_i(t),\ i = 1, \ldots, n$. Then $W(s,t) = \sum_{i=1}^{n} W_i(s,t)$ and $W_i(s,t) \stackrel{i.i.d.}{\sim} \mathrm{WP}(1,\gamma)$. Since $\mathrm{E}W_1(s,t) = \gamma(s,t)$, (a) follows. By (A.2), we have

$$\mathrm{tr}(W_i) = \int_T u_i^2(t)dt = \sum_{r=1}^{\infty} \xi_{ir}^2 \stackrel{d}{=} \sum_{r=1}^{\infty} \lambda_r A_{ir}, \tag{A.3}$$

where A_{ir} are i.i.d., following χ_1^2 for all i and r. Since $\mathrm{tr}(W) = \sum_{i=1}^{n}\mathrm{tr}(W_i)$, (b) follows. (c) follows directly from (b). Noticing that $\mathrm{E}\,\mathrm{tr}(W^{\otimes 2}) = \int_T\int_T \mathrm{E}W^2(s,t)dsdt$ and

$$\begin{aligned}
\mathrm{E}W^2(s,t) &= \mathrm{Var}\,(W(s,t)) + \mathrm{E}^2(W(s,t)) = n\mathrm{Var}\,(W_1(s,t)) + n^2\gamma^2(s,t) \\
&= n\mathrm{E}W_1^2(s,t) - n\mathrm{E}^2(W_1(s,t)) + n^2\gamma^2(s,t) \\
&= n\mathrm{E}u_1^2(s)u_1^2(t) + n(n-1)\gamma^2(s,t),
\end{aligned}$$

we have

$$\begin{aligned}
\mathrm{E}\,\mathrm{tr}(W^{\otimes 2}) &= n\mathrm{E}\left(\int_T u_1^2(t)dt\right)^2 + n(n-1)\,\mathrm{tr}(\gamma^{\otimes 2}) \\
&= n\mathrm{Etr}^2(W_1) + n(n-1)\,\mathrm{tr}(\gamma^{\otimes 2}) \\
&= n(n+1)\,\mathrm{tr}(\gamma^{\otimes 2}) + n\mathrm{tr}^2(\gamma),
\end{aligned}$$

as desired, where we use the result $\mathrm{Etr}^2(W_1) = 2\,\mathrm{tr}(\gamma^{\otimes 2}) + \mathrm{tr}^2(\gamma)$ from Part (c). The proof is complete.

Direct application of Lemma 1 leads to the following useful lemma about unbiased estimators of $\gamma(s,t), \mathrm{tr}(\gamma), \mathrm{tr}^2(\gamma)$, and $\mathrm{tr}(\gamma^{\otimes 2})$.

Lemma 2. *Assume* $W(s,t) \sim \mathrm{WP}(n,\gamma)$ *with* $n > 1$ *and* $\mathrm{tr}(\gamma) < \infty$. *Set* $\hat{\gamma}(s,t) = W(s,t)/n$. *Then* $\hat{\gamma}(s,t)$ *and* $\mathrm{tr}(\hat{\gamma})$ *are the unbiased estimators of* $\gamma(s,t)$ *and* $\mathrm{tr}(\gamma)$ *respectively. Moreover, the unbiased estimators of* $\mathrm{tr}(\gamma^{\otimes 2})$ *and* $\mathrm{tr}^2(\gamma)$ *are, respectively,*

$$\frac{n^2}{(n-1)(n+2)}\left[\mathrm{tr}(\hat{\gamma}^{\otimes 2}) - \frac{1}{n}\mathrm{tr}^2(\hat{\gamma})\right] \ \textit{and}\ \frac{n(n+1)}{(n-1)(n+2)}\left[\mathrm{tr}^2(\hat{\gamma}) - \frac{2}{n+1}\,\mathrm{tr}(\hat{\gamma}^{\otimes 2})\right].$$

B. Proofs of Theorems 1 and 3

Proof of Theorem 1. We only need to verify the two expressions in (2.9), (2.10) follows directly.

To show the first expression of (2.9), notice that we have (2.6). Thus, the q components of $\boldsymbol{w}(t)$ are independent of each other, and the l-th component $w_l(t) \sim \mathrm{GP}(\eta_{wl}, \gamma)$, $l = 1, \ldots, q$, where $\eta_{wl}(t)$ is the l-th component of $\boldsymbol{\eta}_w(t)$. Since $\gamma(s,t)$ has the singular value decomposition (2.7), we have $w_l(t) = \sum_{r=1}^{\infty} \xi_{lr}\phi_r(t)$, where

$$\xi_{lr} = \int_T w_l(t)\phi_r(t)dt \sim N(\mu_{lr}, \lambda_r), \tag{A.4}$$

with $\mu_{lr} = \int_T \eta_{wl}(t)\phi_r(t)dt$. Notice that $\lambda_r > 0$ when $r \le m$ and $\lambda_r = 0$ when $r > m$. It follows that

$$\begin{aligned}\mathrm{ISH} &= \int_T \|\boldsymbol{w}(t)\|^2 dt = \sum_{l=1}^{q} \int_T w_l^2(t)dt = \sum_{l=1}^{q}\sum_{r=1}^{\infty} \xi_{lr}^2 \\ &= \sum_{r=1}^{m}\sum_{l=1}^{q} \xi_{lr}^2 + \sum_{r=m+1}^{\infty}\sum_{l=1}^{q} \xi_{lr}^2,\end{aligned}$$

because the eigenfunctions $\phi_r(t)$ are orthonormal over T. By (A.4), we have $\sum_{r=m+1}^{\infty}\sum_{l=1}^{q} \xi_{lr}^2 = \sum_{r=m+1}^{\infty}\sum_{l=1}^{q} \mu_{lr}^2 = \sum_{r=m+1}^{\infty} \pi_r^2$, and $\sum_{r=1}^{m}\sum_{l=1}^{q} \xi_{lr}^2 \overset{d}{=} \lambda_r A_r$, where $A_r = \sum_{l=1}^{q} \xi_{lr}^2/\lambda_r \sim \chi_q^2(\lambda_r^{-1}\pi_r^2)$ with $\pi_r^2 = \sum_{l=1}^{q} \mu_{lr}^2 = \| \int_T \boldsymbol{\eta}_w(t)\phi_r(t)dt\|^2$. The first expression in (2.9) then follows as desired.

To prove the second expression, recall that $\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t) = (\boldsymbol{I}_n - \boldsymbol{P}_X)\boldsymbol{y}(t) = (\boldsymbol{I}_n - \boldsymbol{P}_X)\boldsymbol{v}(t)$, where $\boldsymbol{v}(t) = [v_1(t), \ldots, v_n(t)]^T$ with $v_i(t) \overset{i.i.d.}{\sim} \mathrm{GP}(0,\gamma)$, and $\boldsymbol{I}_n - \boldsymbol{P}_X = \boldsymbol{U}\,\mathrm{diag}(\boldsymbol{I}_{n-p-1}, \boldsymbol{0}_{(p+1)\times(p+1)})\boldsymbol{U}^T$, where $\boldsymbol{U}$ is an orthonormal matrix of size $n \times n$ so that $\boldsymbol{U}^T\boldsymbol{v}(t) \overset{d}{=} \boldsymbol{v}(t)$. It follows that $W(s,t) = \sum_{i=1}^{n}(y_i(s) - \hat{y}_i(s))(y_i(t) - \hat{y}_i(t)) = \boldsymbol{v}(s)^T(\boldsymbol{I}_n - \boldsymbol{P}_X)\boldsymbol{v}(t) = \sum_{r=1}^{n-p-1} v_r(s)v_r(t) \sim \mathrm{WP}(n-p-1, \gamma)$. Since $\mathrm{ISE} = \int_T W(t,t)dt = \mathrm{tr}(W)$, the second expression in (2.9) follows directly from Lemma 1 (b). Another proof of the second expression can be found in Shen and Faraway (2004). The proof is completed.

Proof of Theorem 3. First notice that under Assumptions A and B and Case (2), $0 < \mathrm{tr}(\gamma) < \infty, \mathrm{tr}(\gamma^{\otimes 2}) < \infty, \delta^2 > 0, \delta_\lambda^2 > 0$, and $0 \le \tau < 1$.

Set $F_n = W_1/W_2$, where W_1 and W_2 denote the numerator and denominator of F_n, respectively, as defined in (2.10). It is easy to see that as $n \to \infty$, we have $W_2 \sim AN\left[\mathrm{tr}(\gamma), 2\,\mathrm{tr}(\gamma^{\otimes 2})/(n-p-1)\right]$. We now show that, for Case (2), W_1 is also asymptotically normally distributed. Under H_{1n}, we have $\pi_r^2 = n^{1-\tau}\delta_r^2$. We can write $W_1 \stackrel{d}{=} (\sum_{r=1}^m \lambda_r A_r + n^{1-\tau} \sum_{r=m+1}^\infty \delta_r^2)/q$, where $A_r \sim \chi_q^2(n^{1-\tau}\lambda_r^{-1}\delta_r^2)$. Notice that

$$\begin{aligned} A_r &\stackrel{d}{=} z_{1r}^2 + \cdots + z_{(q-1)r}^2 + \left[z_{qr} + n^{(1-\tau)/2}\lambda_r^{-1/2}\delta_r\right]^2 \\ &\stackrel{d}{=} A_r^* + 2n^{(1-\tau)/2}\lambda_r^{-1/2}\delta_r z_{qr} + n^{1-\tau}\lambda_r^{-1}\delta_r^2, \end{aligned}$$

where $z_{ir} \stackrel{i.i.d.}{\sim} N(0,1)$ and $A_r^* \sim \chi_q^2$. Thus, we have

$$\sum_{r=1}^m \lambda_r A_r \stackrel{d}{=} \sum_{r=1}^m \lambda_r A_r^* + 2n^{(1-\tau)/2} \sum_{r=1}^m \lambda_r^{1/2}\delta_r z_{qr} + n^{1-\tau} \sum_{r=1}^m \delta_r^2.$$

It follows that $W_1 \stackrel{d}{=} \sum_{r=1}^m \lambda_r A_r^*/q + 2n^{(1-\tau)/2} \sum_{r=1}^m \lambda_r^{1/2}\delta_r z_{qr}/q + n^{1-\tau}\delta^2/q$. Since $\delta^2 > 0$ and $\delta_\lambda^2 > 0$, as $n \to \infty$, the last two terms on the right-hand side dominate the first term. Therefore, W_1 is asymptotically normally distributed, i.e., $W_1 \sim AN\left[n^{1-\tau}\delta^2/q, 4n^{1-\tau}\delta_\lambda^2/q^2\right]$, by dropping higher order terms. It follows that $F_n = W_1/W_2$ is also asymptotically normally distributed with

$$\begin{aligned} \mathrm{E}(F_n) &= \frac{\mathrm{E}(W_1)}{\mathrm{E}(W_2)}[1+o(1)] = \frac{n^{1-\tau}\delta^2}{q\,\mathrm{tr}(\gamma)}[1+o(1)], \\ \mathrm{Var}\,(F_n) &= \left[\frac{1}{\mathrm{E}^2(W_2)}\mathrm{Var}\,(W_1) + \frac{\mathrm{E}^2(W_1)}{\mathrm{E}^4(W_2)}\mathrm{Var}\,(W_2)\right][1+o(1)] \\ &= \frac{4n^{1-\tau}\delta_\lambda^2}{q^2\mathrm{tr}^2(\gamma)}[1+o(1)], \end{aligned}$$

dropping the higher order terms. The proof of (2.14) is then complete.

To prove (2.15), notice that

$$\begin{aligned} &P\left(F_n \geq F_{n,\alpha}^* | H_{1n}\right) = P\Big(\frac{F_n - E(F_n)}{\sqrt{\mathrm{Var}\,(F_n)}} \geq \frac{F_{n,\alpha}^* - \mathrm{E}(F_n)}{\sqrt{\mathrm{Var}\,(F_n)}} | H_{1n}\Big) \\ &= 1 - \Phi\Big(\frac{F_{n,\alpha}^* - \mathrm{E}(F_n)}{\sqrt{\mathrm{Var}\,(F_n)}}\Big) + o(1) = 1 - \Phi\Big(\frac{[T_\alpha^* - n^{1-\tau}\delta^2]/[q\,\mathrm{tr}(\gamma)]}{\sqrt{4n^{1-\tau}\delta_\lambda^2/[q\,\mathrm{tr}(\gamma)]^2}}\Big) + o(1) \\ &= \Phi\Big(\frac{n^{(1-\tau)/2}\delta^2}{2\delta_\lambda}\Big) + o(1), \end{aligned}$$

which obviously tends to 1 as $n \to \infty$. The theorem is proved.

References

Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika* **51**, 285-311.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing* Marcel Dekker, New York.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications* Chapman and Hall, London.

Fan, J. and Lin, S. K. (1998). Test 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.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall, London.

Hall, P. and van Keilegom, I. V. (2007). Two-sample tests in functional data analysis starting from discrete data. *Statist. Sinica* **17**, 1511-1531.

Huynh, H. and Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *J. Educ. Statist.* **1**, 69-82.

Ramsay, J. O. (1995). Some tools for the multivariate analysis of functional data. In *Recent Advances in Descriptive Multivariate Analysis* (Edited by W. Krzanowski), 269-282. Clarendon, Oxford.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis (with discussion). *J. Roy. Statist. Soc. Ser. B* **53**, 539-572.

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

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

Ruppert, D., Wand, M. P and Carroll, R. J. (2003). *Semiparametric Regression.* Cambridge University Press.

Shen, Q. and Faraway, J. (2004). An F test for linear models with functional responses. *Statist. Sinica* **14**, 1239-1257.

Shen, Q. and Xu, H. (2007). Diagnostics for linear models with functional responses. *Technometrics* **49**, 26-33.

Wahba, G. (1990). Spline models for observational data. SIAM, Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. **59**.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

Yang, X., Shen, Q., Xu, H. and Shoptaw, S. (2007). Functional regression analysis using an F test for longitudinal data with large numbers of repeated measures. *Statist. Medicine* **26**, 1552-1566.

Zhang, J. T. (2005). Approximate and asymptotic distributions of χ^2-type mixtures with applications. *J. Amer. Statist. Assoc.* **100**, 273-285.

Zhang, J. T. and Chen, J. W. (2007). Statistical inferences for functional data. *Ann. Statistics* **35**, 1052-1079.

Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546.

E-mail: stazjt@nus.edu.sg

(Received November 2009; accepted April 2010)