Statistics in Medicine

Research Article

Received 17 December 2010

Accepted 3 June 2011

Published online 22 July 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4316

A quasi *F*-test for functional linear models with functional covariates and its application to longitudinal data

Hongquan Xu,^{a*†} Qing Shen,^b Xiaowei Yang^{c,d} and Steven Shoptaw^e

Functional linear models are useful in analyzing data from designed experiments and observational studies with functional responses, as well as longitudinal data with a large number of repeated measures on each subject. We propose a quasi F-test for functional linear models with functional covariates and outcomes. We develop a numerical procedure and an efficient approximation for computing p-values, and present a simple way to test individual predictors. For illustration, we apply the proposed procedure to a longitudinal depression data set with repeatedly measured methamphetamine use as a predictor. We conduct a simulation study to assess the size and the power of the test. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: chi-squared approximation; *F*-test; functional data analysis; functional regression analysis; longitudinal data analysis

1. Introduction

In biomedical research with longitudinal designs, subjects are repeatedly measured for a set of characteristics so that time-varying relationships between the responses and explanatory variables of interest are modeled [1–3]. In certain fields of study, such as substance abuse, environmental, and public health research, repeated measures are sometimes collected at rapid frequencies over long periods of time. For example, in a 12-week smoking cessation clinical trial, the investigators collected up to 36 breath samples on each smoker and analyzed them for testing the effectiveness of two behavioral therapies [4]. In a 16-week depression and methamphetamine abuse study, the investigators collected up to 16 depression evaluations and 48 urine samples on each subject to compare outcomes from four treatment conditions [5]. In such studies, the investigators included frequent repeated measures to describe the process of outcomes to experimental treatments in real time, which can be particularly vital to the study of treatments for chronic diseases with high rates of relapse.

Linear models are frequently used for interpreting or predicting responses by a set of covariates or predictors. Many authors have studied the form of functional linear models: $\mathbf{Y}(t) = \mathbf{X}\boldsymbol{\beta}(t) + \boldsymbol{\epsilon}(t)$, where the responses are functions and the covariates are scalar vectors (see, for example, [6–13]). Ramsay and Silverman [6] laid out some general ideas on estimation and provided preliminary methods for inference. Faraway [7] pointed out the inappropriateness of traditional multivariate test statistics and proposed a bootstrap-based testing method. Fan and Lin [8] proposed adaptive transform-based tests for functional analysis of variance models. Shen and Faraway [9] proposed a functional *F*-test for comparing nested functional linear models, and Shen and Xu [10] considered diagnostics for such models. Yang *et al.* [11]

^aDepartment of Statistics, University of California, Los Angeles, CA 90095, USA

^bEdmunds.com Inc., 1620 26th St, Suite 400 South, Santa Monica, CA 90404, USA

^cDivision of Biostatistics, Department of Public Health Sciences, School of Medicine, University of California, Davis, CA 95616, USA

^dBayessoft, Inc., 2221 Caravaggio Drive, Davis, CA 95618, USA

^eDepartment of Family Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

^{*}Correspondence to: Hongquan Xu, Department of Statistics, University of California, Los Angeles, CA 90095, USA. †E-mail: hqxu@stat.ucla.edu

and Yang and Nie [12] used such models to analyze a longitudinal data set from a clinical trial. Zhang and Chen [13] proposed an L_2 -norm-based test statistic and studied its asymptotic properties.

In many real-life research settings, however, both responses and predictors can be functions. Thus, the problem of model fitting and making statistical inference for functional linear models $\mathbf{Y}(t) = \mathbf{X}(t)\boldsymbol{\beta}(t) + \boldsymbol{\epsilon}(t)$ becomes increasingly interesting. Brumback and Rice [14] and Hoover *et al.* [15] used such models for longitudinal data and assumed that the functional coefficients $\boldsymbol{\beta}(t)$ are smooth. Such models, referred to as concurrent or pointwise models by Ramsay and Silverman [6], belong to the class of varying coefficient models [16]. Wu and Yu [17] and Fan and Zhang [18] reviewed various methods involving smoothing techniques for estimation and inference with varying coefficient models.

We take a simple and direct approach for estimation and inference for functional linear models, where both the response and the covariates are functions or values sampled from continuous spaces of the functions. Following prior examples [6, 7, 9], we estimate the parameters by pointwise least squares. In practice, when we collect data at fixed time points, our pointwise least squares estimator is a nonsmoothed and unbiased estimator of $\beta(t)$. If desirable, one can use local kernel regression or other smoothing techniques to get a smoothed and hopefully more efficient estimator of $\beta(t)$. However, one must be aware of the problem of oversmoothing. Faraway [7] gave a good discussion on this issue.

In this paper, we focus on the inference of $\boldsymbol{\beta}(t)$, which is known to be a challenging problem. Traditional multivariate test statistics cannot be used when covariates are functions. A naive approach is to examine the pointwise t or F-statistics on each time point for testing $\boldsymbol{\beta}(t)$. This carries a serious problem with multiple comparisons, and if we apply Bonferroni corrections to the significance level, power would be significantly compromised because responses are often highly correlated within each subject. We propose a quasi F-statistic for testing nested models, which is an extension of the so-called *functional* F-statistic of Shen and Faraway [9] for functional response with scalar covariates. The quasi F-statistic takes the same form as the ordinary scalar F-statistic: $QF = \frac{(rss_{\omega} - rss_{\Omega})/(p-q)}{rss_{\Omega}/(n-p)}$, where rss_{ω} and rss_{Ω} are the residual sums of squares under nested models ω and Ω with q and p covariates (p > q), respectively. We use the term 'quasi' to emphasize the fact that the numerator and denominator of the quasi F-statistic are not independent when the covariates change with time, whereas they are independent in the scalar or functional F-statistic.

The exact distribution of the quasi F (QF) statistic is rather complicated and does not seem to have a simple close form. One possible approach is to use the bootstrap method [7]. Assuming that the errors are from a Gaussian stochastic process, we can generate many samples and use those simulated samples to approximate the distribution of the QF statistic. This approach is general but requires intensive computation. Here, we propose a numerical procedure to compute the null distribution exactly. We further propose a χ^2 approximation method for computing *p*-values. The approximation requires much less computation than both the bootstrap method and the numerical procedure. The simulation studies we conducted show that the approximation is fast and accurate for practical use.

A linear model often includes many covariates, and it is desirable to test the significance of individual covariates. One advantage of the quasi F-test is that we can compute directly the QF statistic for any single covariate from the full model. Indeed, the QF statistic is a weighted average of the pointwise F-statistics, and the weights are proportional to the pointwise estimated variances. This provides a convenient way to conduct variable or model selection in practice.

In Section 2, we present the underlying distribution theory for the quasi F-statistic and provide a practical procedure for conducting the quasi F-test. Not surprisingly, when covariates do not vary with time, the quasi F-test is identical to the functional F-test that Shen and Faraway [9] proposed. In Section 3, we apply the proposed quasi F-test to a clinical trial [5] with functional responses and functional covariates. In Section 4, we conduct simulations to study the size and the power of the quasi F-test and compare it with a linear mixed effects model approach. Finally, Section 5 provides concluding remarks.

2. A quasi F-test for functional linear models

Given a functional response $y_i(t)$ and a vector of predictors $\mathbf{x}_i(t) = (x_{i1}(t), \dots, x_{ip}(t))^T$ for $i = 1, \dots n$, the functional regression model takes the familiar form

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

)

where $\boldsymbol{\beta}(t) = (\beta_1(t), \dots, \beta_p(t))^T$ are unknown coefficient functions and $\epsilon_i(t)$ is a Gaussian stochastic process with mean zero and covariance function $\gamma(s, t)$. Here, we assume $\epsilon_i(\cdot)$ and $\epsilon_j(\cdot)$ to be independent for $i \neq j$.

The pointwise least squares estimator of $\boldsymbol{\beta}(t)$ is $\hat{\boldsymbol{\beta}}(t) = (\mathbf{X}(t)^T \mathbf{X}(t))^{-1} \mathbf{X}(t)^T \mathbf{Y}(t)$, where $\mathbf{X}(t) = (\mathbf{x}_1(t), \dots, \mathbf{x}_n(t))^T$ is the usual $n \times p$ design matrix and $\mathbf{Y}(t) = (y_1(t), \dots, y_n(t))^T$ is a vector of responses. The predicted responses are $\hat{y}_i(t) = \mathbf{x}_i(t)^T \hat{\boldsymbol{\beta}}(t)$ with residuals $\hat{\epsilon}_i(t) = y_i(t) - \hat{y}_i(t)$ and residual sum of squares $rss = \sum_{i=1}^n \int \hat{\epsilon}_i(t)^2 dt$.

In practice, we usually do not observe $y_i(t)$ for all t but only $y_i(t_{ij})$, where $(t_{i1}, ..., t_{im})$ is the sampling by time grid for subject i. It is desirable to collect data at the same time grid, $t_1, ..., t_m$, for easy interpretation and estimation and is common to many well-designed experiments or studies. It is frequently the case that the time grid varies between subjects because of such factors as early termination and loss to follow-up. In such instances, by assuming an ignorable missingness mechanism, we can use smoothing technique to get fixed time points [6, 7]. Within the scope of this paper, we assume that the responses are observed on evenly spaced, fixed time points t_1, \ldots, t_m . Then, the functional model in Equation (1) becomes

$$y_i(t_j) = \mathbf{x}_i(t_j)^T \boldsymbol{\beta}(t_j) + \epsilon_i(t_j), \text{ for } i = 1, \dots, n; j = 1, \dots, m,$$
(2)

and we can do pointwise estimation and regression. The random errors are independent between subjects but normally distributed within each subject with mean zero and covariance matrix Σ . We can replace the integration with summation and compute $rss = \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{\epsilon}_i(t_j)^2$. For the rest of the paper, we consider the multivariate formulation in Equation (2) rather than the usually unobservable functional model (1).

An important inference problem is to test whether we can simplify a linear model. Given the model Ω in Equation (2) with *p* covariates, we want to test whether a smaller model ω with *q* covariates is sufficient. Model ω typically consists of a subset of predictors in Ω , but it can be a linear subspace of Ω in general. Model Ω is often referred to as the full model, and ω is the null or reduced model. In the spirit of Shen and Faraway [9], we define a quasi *F*-statistic,

$$QF = \frac{(rss_{\omega} - rss_{\Omega})/(p-q)}{rss_{\Omega}/(n-p)},$$
(3)

where rss_{ω} and rss_{Ω} are residual sums of squares under models ω and Ω , respectively. Under the null model ω , both rss_{Ω} and $rss_{\omega} - rss_{\Omega}$ are quadratic forms in normal random variables; therefore, they are distributed like linear combinations of independent χ^2 random variables.

When the covariates do not vary with t (i.e., $\mathbf{x}_i(t) = \mathbf{x}_i$ are fixed), $\operatorname{rss}_{\Omega}$ and $\operatorname{rss}_{\omega} - \operatorname{rss}_{\Omega}$ are independent of each other. Using the Satterthwaite [19] approximation, Shen and Faraway [9] approximated the QF distribution by an ordinary scalar F-distribution with degrees of freedom (DOF) $df_1 = \lambda(p-q)$ and $df_2 = \lambda(n-p)$, where $\lambda = \operatorname{trace}(\Sigma)^2/\operatorname{trace}(\Sigma^2)$ is called the degrees-of-freedom-adjustment-factor.

When the covariates vary with t, rss_{Ω} and $rss_{\omega} - rss_{\Omega}$ are no longer independent, and we cannot approximate the QF distribution by an ordinary scalar F-distribution. Nevertheless, we can compute the null distribution numerically by using some well-established results on quadratic forms. Specifically, for an observed QF value $f_0 > 0$, we can compute the p-value as

$$P(QF \ge f_0) = P\left(\frac{rss_{\omega}}{rss_{\Omega}} \ge r_0\right) = P(rss_{\omega} - r_0 rss_{\Omega} \ge 0), \tag{4}$$

where $r_0 = 1 + f_0(p-q)/(n-p)$ is the observed value of ratio $rss_{\omega}/rss_{\Omega}$. In the last expression, $rss_{\omega} - r_0 rss_{\Omega}$ is a quadratic form in normal random variables and is distributed like a linear combination of independent χ^2 random variables, $\sum_{j=1}^k \lambda_j \chi_j^2(1)$, where all $\chi_j^2(1)$'s are χ^2 random variables with 1 DOF and are independent of each other and $\lambda_1, \ldots, \lambda_k$ are nonzero eigenvalues of an $(nm) \times (nm)$ matrix **A**, which is a function of r_0 , covariance matrix Σ and design matrices $\mathbf{X}(t_j)$; see Theorem 2 in the following text for precise definition. Given $\lambda_1, \ldots, \lambda_k$, efficient algorithms are available in the literature to compute the probability exactly [20, 21].

One disadvantage to this approach is that obtaining the eigenvalues of an $(nm) \times (nm)$ matrix is computationally expensive because nm is often very large for functional data. A practical approach is to use some approximation. It is well studied in the literature that a linear combination of independent χ^2 random variables can be efficiently approximated by a single χ^2 random variable. The widely used Satterthwaite method approximates $\sum_{j=1}^{k} \lambda_j \chi_j^2(1)$ with $a\chi^2(d)$ by matching the first two moments, where *d* is the DOF. However, the approximation is poor because we have both positive and negative λ_j 's. Here, we adopt the Pearson three-moment approximation [20, 22]. Specifically, we approximate $\sum_{j=1}^{k} \lambda_j \chi_j^2(1)$ with $a\chi^2(d) + b$ by matching the first three moments. According to [20],

$$d = \frac{c_2^3}{c_3^2}, a = \frac{c_3}{c_2}, b = c_1 - \frac{c_2^2}{c_3}$$

where $c_i = \sum_{j=1}^{k} \lambda_j^i$ for i = 1, 2, 3. For accuracy, the approximate DOF, d, are not necessarily an integer. This is not a problem in practice and most software (including R) can handle fractional DOF. Then, we can approximate the probability in Equation (4) as

$$P(rss_{\omega} - r_0 rss_{\Omega} \ge 0) \approx P(a\chi^2(d) + b \ge 0) = \begin{cases} P(\chi^2(d) \ge q_0) & \text{if } a > 0\\ P(\chi^2(d) \le q_0) & \text{if } a < 0, \end{cases}$$
(5)

where $q_0 = -b/a = c_2(c_2^2 - c_1c_3)/c_3^2$. The approximation avoids the time-consuming eigenvalue decomposition explicitly because $c_i = \text{trace}(\mathbf{A}^i)$. Imhof [20] compared the approximation with exact numerical computation via integration for several quadratic forms and concluded that the approximation is sufficient for certain practical purposes. Zhang [23] provided a theoretical justification by establishing an upper bound on the approximation error. Our computation and simulation further supported this conclusion, and therefore we recommend the use of approximation in practice for calculating *p*-values.

The distribution of the QF statistic in Equation (3) depends on the eigenvalues of matrix **A**, which is a function of the covariance matrix Σ that is often unknown to the data analyst. This does not raise a problem because we can estimate it by the empirical covariance matrix

$$\hat{\boldsymbol{\Sigma}} = \left[\frac{1}{n-p} \sum_{i=1}^{n} \hat{\epsilon}_i(t_j) \hat{\epsilon}_i(t_k)\right]_{m \times m}$$
(6)

under the full model Ω , where j, k = 1, ..., m. Large DOF, say $n - p \ge 30$, are desirable for an accurate estimation.

The proposed QF statistic is scale free in the sense that we get the same sampling distribution if we replace Σ with $c\Sigma$ for any constant c > 0. In particular, we get the same result if the coefficient 1/(n-p) in Equation (6) is replaced with any constant. This is an important and desirable property because we need to estimate Σ from data in practice.

The following theorem shows the relationship between the QF statistic and the pointwise F-statistics.

Theorem 1

The QF statistic in Equation (3) equals to

$$QF = \frac{\sum_{j=1}^{m} F_j \hat{\sigma}_j^2}{rss_{\Omega}/(n-p)}$$

where F_j and $\hat{\sigma}_j^2$ are pointwise *F*-statistic and the estimated variance at time point t_j .

It can be shown (see the proof of Theorem 1 given in the Appendix) that $\sum_{j=1}^{m} \hat{\sigma}_{j}^{2} = rss_{\Omega}/(n-p)$. Therefore, the QF statistic is simply a weighted average of the pointwise *F*-statistics, and the weights are proportional to the pointwise estimated variances.

When fitting linear models, it is often interesting to know which covariates are important to be included. To test whether a particular covariate is significant, we can fit a model without that covariate and perform a quasi *F*-test. It can be very tedious to repeat this procedure for every predictor. Fortunately, Theorem 1 shows that we can compute directly the QF statistics from the full model. Specifically, to test whether the *l*th covariate is significant, (i.e., whether $\beta_l(t) = 0$), we can obtain $F_j \hat{\sigma}_j^2 = \hat{\beta}_l^2(t_j)/(\mathbf{X}(t_j)^T \mathbf{X}(t_j))_{ll}^{-1}$ directly from the full model, where $(\mathbf{X}(t_j)^T \mathbf{X}(t_j))_{ll}^{-1}$ is the *l*th diagonal element of $(\mathbf{X}(t_j)^T \mathbf{X}(t_j))^{-1}$ for l = 1, ..., p. This provides a convenient way for variable selection.

The rest of this section provides technical details of the underlying theory and computation issues. Readers who are interested in applications can go to the next section directly.

Let $\mathbf{E} = (\epsilon_i(t_j))$ be the $n \times m$ matrix of measurement errors. Denote the rows of \mathbf{E} as $\epsilon_1^T, \ldots, \epsilon_n^T$. By the assumption, we have $\epsilon_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\operatorname{cov}(\epsilon_i, \epsilon_j) = \mathbf{0}$ for $i \neq j$. Denote the columns of \mathbf{E} as $\epsilon_{\cdot t_1}, \ldots, \epsilon_{\cdot t_m}$ and let vec(E) be the $nm \times 1$ vector that consists of all elements of E stacked by column. Then, $\operatorname{cov}(\operatorname{vec}(E)) = \Sigma \bigotimes I_n = (\sigma_{ij}I_n)$, which is the Kronecker product of Σ and the $n \times n$ identity matrix I_n .

Let $\mathbf{H}(t_j) = \mathbf{X}(t_j)(\mathbf{X}(t_j)^T \mathbf{X}(t_j))^{-1} \mathbf{X}(t_j)^T$ be the $n \times n$ hat matrix. Following the scalar univariate linear model theory [24–26], $\hat{\boldsymbol{\epsilon}}_{\cdot t_j} = (\mathbf{I} - \mathbf{H}(t_j))\boldsymbol{\epsilon}_{\cdot t_j}$ are the residual vectors and $\mathbf{I} - \mathbf{H}(t_j)$ are idempotent. Then,

$$rss = \sum_{j=1}^{m} \hat{\boldsymbol{\epsilon}}_{\cdot t_{j}}^{T} \hat{\boldsymbol{\epsilon}}_{\cdot t_{j}} = \sum_{j=1}^{m} \boldsymbol{\epsilon}_{\cdot t_{j}}^{T} (\mathbf{I} - \mathbf{H}(t_{j})) \boldsymbol{\epsilon}_{\cdot t_{j}} = \operatorname{vec}(\mathbf{E})^{T} \mathbf{M} \operatorname{vec}(\mathbf{E}),$$

where $\mathbf{M} = \text{diag}(\mathbf{I} - \mathbf{H}(t_1), \dots, \mathbf{I} - \mathbf{H}(t_m))$ is an $(nm) \times (nm)$ block diagonal matrix. It is clear now that *rss* is a quadratic form in normal random variables.

For the given nested linear models ω and Ω , we denote the corresponding **M** matrices as \mathbf{M}_{ω} and \mathbf{M}_{Ω} . Then, under model ω , $rss_{\omega} - r_0 rss_{\Omega} = \text{vec}(\mathbf{E})^T (\mathbf{M}_{\omega} - r_0 \mathbf{M}_{\Omega}) \text{vec}(\mathbf{E})$. Applying Theorem 2.1 of [27], we have the following result.

Theorem 2

Under the null model ω , $rss_{\omega} - r_0 rss_{\Omega}$ is distributed like a linear combination of independent χ^2 random variables, that is,

$$rss_{\omega} - r_0 rss_{\Omega} \sim \sum_{j=1}^k \lambda_j \chi_j^2(1),$$

where each χ^2 random variable has 1 DOF and is distributed independently of each other and $\lambda_1, \ldots, \lambda_k$ are all real nonzero eigenvalues of $\mathbf{A} = (\boldsymbol{\Sigma} \bigotimes \mathbf{I}_n)(\mathbf{M}_{\omega} - r_0 \mathbf{M}_{\Omega})$.

The matrix **A** in Theorem 2 is not symmetric. Nevertheless, all the eigenvalues of **A** are real because $\Sigma \bigotimes I_n$ is positive semidefinite and $M_{\omega} - r_0 M_{\Omega}$ is symmetric.

Now, we describe a procedure for computing *p*-values for the quasi *F*-test.

- (1) Perform pointwise least squares estimation for both models ω and Ω .
- (2) Compute the QF statistic as in Equation (3); denote it as f_0 and let $r_0 = 1 + f_0(p-q)/(n-p)$.
- (3) Estimate Σ by $\hat{\Sigma}$ in Equation (6) under the full model Ω .
- (4) Generate matrix $\hat{\mathbf{A}} = (\hat{\boldsymbol{\Sigma}} \bigotimes \mathbf{I}_n)(\mathbf{M}_{\omega} r_0 \mathbf{M}_{\Omega}).$
- (5) Compute $c_1 = \operatorname{trace}(\hat{\mathbf{A}}), c_2 = \operatorname{trace}(\hat{\mathbf{A}}^2), c_3 = \operatorname{trace}(\hat{\mathbf{A}}^3), d = c_2^3/c_3^2, a = c_3/c_2$, and $q_0 = c_2(c_2^2 c_1c_3)/c_3^2$. Or alternatively, find the eigenvalues of $\hat{\mathbf{A}}$ and denote the nonzero eigenvalues as $\hat{\lambda}_1, \dots, \hat{\lambda}_k$.
- (6) Compute $p = P(\chi^2(d) \ge q_0)$ for a > 0 or $p = P(\chi^2(d) \le q_0)$ for a < 0. Or alternatively, use Davis's algorithm [21] to evaluate $p^* = P(\sum_{j=1}^k \hat{\lambda}_j \chi_j^2(1) > 0)$ numerically.

We report p or p^* as the p-value for the observed value f_0 . The main computational cost here is the eigenvalue decomposition in step 5. Davis's algorithm is quite efficient even for k as large as 1500. When nm is not very large, say < 1000, we can compute the p^* value using Davis's algorithm. For larger nm, we can use p to approximate p^* . As our computation suggests, the three-moment χ^2 approximation is satisfactory for assessing significance; therefore, we recommend to use p as the p-value in practice.

We can use the procedure to report the significance of individual covariates in a model after some minor modifications. We fit the model once and compute the QF statistics as in Theorem 1. We need to estimate the covariance matrix and compute matrices $\hat{\Sigma} \otimes I_n$ and M_{Ω} once. However, we need to compute M_{ω} for each model that includes all but one covariates.

We need to modify the procedure properly to deal with missing values, which are common in longitudinal data. In step 1, we perform least squares estimation only to the observed data at each time point, which yields different DOF for different time points. Therefore, we replace n - p by the average DOF for model Ω in step 2. Step 3 is the most crucial step for which we propose two methods for estimating $\Sigma = (\sigma_{ij})$. The first method uses subjects with complete residuals (i.e., missing values are handled by complete case analysis). This may end up with discarding too many cases and result in an inaccurate estimation. The second method estimates σ_{ij} using all available pairs of residuals in columns *i* and *j*. This has the advantage of using more observations in the estimation of the covariance but can result in a matrix, which may not be positive semidefinite. An alternative solution is to apply the method of multiple imputation with smoothing techniques to enforce complete data sets. In step 4, we simply delete rows and columns in $\hat{\Sigma} \otimes I_n$ that correspond to missing residuals.

When the covariates do not vary over time t, the computation is much easier. Note that in this case the hat matrices $\mathbf{H}(t_j) = \mathbf{H}$ are the same, so $\mathbf{M} = \mathbf{I}_m \bigotimes (\mathbf{I} - \mathbf{H})$. The properties of the Kronecker product imply that $(\Sigma \bigotimes \mathbf{I}_n)\mathbf{M} = (\Sigma \bigotimes \mathbf{I}_n)(\mathbf{I}_m \bigotimes (\mathbf{I} - \mathbf{H})) = \Sigma \bigotimes (\mathbf{I} - \mathbf{H})$, and the eigenvalues of $\Sigma \bigotimes (\mathbf{I} - \mathbf{H})$ are simply the products of eigenvalues of Σ and $\mathbf{I} - \mathbf{H}$ [26]. Note that $\mathbf{I} - \mathbf{H}$ is idempotent, and its eigenvalues are either 0 or 1. Consequently, the eigenvalues of $\Sigma \bigotimes (\mathbf{I} - \mathbf{H})$ are completely determined by the eigenvalues of Σ and the rank of $\mathbf{I} - \mathbf{H}$. Then, $rss_{\omega} - rss_{\Omega} \sim \sum_{j=1}^{m} \lambda_j \chi_j^2 (p-q)$ and $rss_{\Omega} \sim \sum_{j=1}^{m} \lambda_j \chi_j^2 (n-p)$, where $\lambda_1 \ge \ldots \ge \lambda_m \ge 0$ are the eigenvalues of Σ . Furthermore, they are independent because $\mathbf{M}_{\Omega}(\Sigma \bigotimes \mathbf{I}_n)(\mathbf{M}_{\omega} - \mathbf{M}_{\Omega}) = \Sigma \bigotimes (\mathbf{I} - \mathbf{H}_{\Omega})(\mathbf{H}_{\Omega} - \mathbf{H}_{\omega}) = 0$, where the last equation follows the scalar univariate linear model theory [26]. Therefore, the *p*-value is $P(\sum_{j=1}^{m} \lambda_j \chi_j^2 (p-q) - (r_0 - 1) \sum_{j=1}^{m} \lambda_j \chi_j^2 (n-p) > 0)$. Alternatively, $(\Sigma \bigotimes \mathbf{I}_n)(\mathbf{M}_{\omega} - r_0\mathbf{M}_{\Omega}) =$ $\Sigma \bigotimes ((\mathbf{I} - \mathbf{H}_{\omega}) - r_0(\mathbf{I} - \mathbf{H}_{\Omega}))$. It is easy to show that the distinct eigenvalues of $(\mathbf{I} - \mathbf{H}_{\omega}) - r_0(\mathbf{I} - \mathbf{H}_{\Omega})$ are $1 - r_0$, 1 and 0, with multiplicity n - p, p - q, and q, respectively. This leads to the same formula for the *p*-value. For computing the *p*-value, we need to find the eigenvalues of the $m \times m$ matrix $\hat{\Sigma}$ only, which is an easy task. So, we can compute the *p*-value of the functional *F*-test of Shen and Faraway [9] efficiently if requested.

3. Application

We applied the proposed quasi F-procedure to a randomized clinical trial [5], which is a 16-week study conducted in outpatient treatment research clinics in the Hollywood and the West Hollywood areas of Los Angeles from 1996 to 2001. All the participants were diagnosed as methamphetamine dependent, were self-identified gay or bisexual men, and were seeking treatment for their dependence on methamphetamine. All the participants began with a 2-week baseline period, and 162 subjects completing the baseline periods were randomly assigned to one of four treatment conditions: cognitive behavioral therapy (CBT, $N_1 = 40$), gay-specific CBT (GCBT; $N_2 = 40$), contingency management (CM; $N_3 = 42$), and combined CBT and CM (CBT + CM; $N_4 = 40$). CBT was a standard intervention control condition, and the CM procedures consisted of providing vouchers of increasing value for urine samples documenting continuous abstinence from methamphetamine use. During the 16-week study, methamphetamine use was measured using thrice-weekly urine samples analyzed for drug metabolite. Self-reported depressive symptoms were collected weekly using the Beck Depression Inventory (BDI). Mild depression was indicated by scores of 10-19, moderate depression was indicated by scores of 19-29, and BDI totals over 30 indicated severe depression. Initial BDI scores, self-report of recent methamphetamine use, and HIV status were of the most interest and collected before randomization. Other demographical variables were also collected at baseline, but they were less important to the investigators. There were substantial missing observations during the trial because of the subjects' missing clinical visits or nonresponse to certain questions on the self-evaluation forms. The missingness became more severe as the study continued. Only 11% of the BDI scores were missing at week 1, but the number was doubled at week 2 and grew to 52% at week 15. Only 32 subjects (20%) had complete records on covariates and repeated measures. In this paper, we assumed that missing values are ignorable in the sense that we can use the observed values to obtain unbiased estimates.

Figure 1 shows the average BDI scores by treatments at baseline (week 0) and during treatments (weeks 1–16). It is suggestive that all treatments were associated with reduction of depression level in the first 2 to 3 weeks, with treatments CM and CBT + CM appearing more effective than the other two. The questions of interest for our analysis were the following: (i) whether there are any significant differences of depression improvement across treatment conditions as measured by BDI scores; (ii) whether depression scores are associated with methamphetamine use; and (iii) whether depression scores depend on HIV status.

We defined a variable, Drug, to represent whether a subject used methamphetamine during a week as follows: Drug = 1 if at least one of the three urine tests were positive and Drug = 0 if none of the tests were positive. Drug was missing if a subject missed all three urine tests in a week. Note that Drug is a functional covariate, that is, time-varying covariate. The distribution of BDI scores looked positively skewed, and we applied a square root transformation to make it approximately normal. For each subject,



Figure 1. Average Beck Depression Inventory (BDI) scores by treatments at baseline (week 0) and during treatment (weeks 1–16). CBT, cognitive behavioral therapy; GCBT, gay-specific cognitive behavioral therapy; CM, contingency management.

we fitted the following initial (full) model, which includes all the variables of interest such as baseline BDI scores, *Drug*, HIV status, and treatments:

$$y_i(t) = \beta_0(t) + BaseScore_i \cdot \beta_1(t) + Drug_i(t) \cdot \beta_2(t) + HIV_i \cdot \beta_3(t) + CBT_i \cdot \beta_4(t) + GCBT_i \cdot \beta_5(t) + CM_i \cdot \beta_6(t) + \epsilon_i(t),$$
(7)

where $y_i(t)$ is the square root of BDI score at week t on subject i = 1, ..., 162 and $BaseScore_i$ is the square root of BDI score at week 0. $Drug_i(t)$ is 1 if the subject used methamphetamine during week t and 0 otherwise. HIV_i is 1 if the subject was HIV positive and 0 otherwise; CBT_i , $GCBT_i$, and CM_i are indicators of the treatment the subject received (treatment CBT + CM serves as a reference).

Figure 2 displays the pointwise estimates of the covariate coefficients $\beta_l(t)$ s and their corresponding 95% CIs. Because of missing values in the response and drug use, these pointwise models have different DOF, which vary from 65 (week 15) to 130 (week 1) and have an average of 91. We found that *BaseScore* have the most significant impact on depression during the study: all the estimates of $\beta_1(t)(t = 1, ..., 16)$ are positive, and 13 of the 16 CIs exclude 0. The effect of *BaseScore* decreased as the treatments progressed. For *Drug*, all the estimates are positive, and 6 of the CIs exclude 0. For *HIV*, none of the CIs exclude 0, although all estimates are positive. For *CBT* and *GCBT*, 4 and 2 CIs exclude 0, respectively. For *CM*, no CIs exclude 0. The pointwise estimates and the CIs provide vivid interpretations regarding how the effects changed with time; however, it is not clear whether covariates have overall significance by summarizing over the 16 time points.

The theory and procedure in the previous section provide an efficient way of testing the significance of covariates. We used the pairwise available residuals to estimate Σ . The results looked similar when we used only the complete residuals, and we did not present these here. Table I shows the observed statistics f_0 , associated χ^2 approximation parameters $(a, d, \text{ and } q_0)$ and their *p*-values using the approximation and numerical integration. The approximation is satisfactory, and both *p*-values lead to the same conclusions. *BaseScore* and *Drug* are very significant predictors (p < .01) but *HIV* is not. Both *CBT* and *GCBT* are significant predictors of y(t) at the 10% level, that is, treatment effects of CBT and GCBT are significantly different from that of the treatment of CBT + CM. However, *CM* is not a significant predictor at the 10% level, implying that CM and CBT + CM are similarly effective in reducing depression scores. These findings are consistent with Figure 1, which shows that treatments CBT and GCBT are similarly effective but different from CBT + CM, which is similarly effective to CM. Note that the latter two treatments provided financial incentives for being methamphetamine abstinent, whereas the former two did not. It is of special interest to examine the effect of financial incentives used by the CM method by comparing the first two treatments with the last two. For this purpose, we created a new variable, *Incentive*, and considered the following model:

$$y_i(t) = \beta_0(t) + BaseScore_i \cdot \beta_1(t) + Drug_i(t) \cdot \beta_2(t) + Incentive_i \cdot \beta_3(t) + \epsilon_i(t), \quad (8)$$

where $Incentive_i = 1$ if subject *i* was assigned treatment CM or CBT + CM and $Incentive_i = 0$ otherwise. Figure 3 shows the pointwise estimates and 95% CIs of the *Incentive* effect (i.e., $\beta_3(t)$). All the estimates are negative, and 4 CIs exclude 0, implying that offering incentives reduced depression scores.



Figure 2. Pointwise estimates and 95% confidence intervals in the initial model. CBT, cognitive behavioral therapy; GCBT, gay-specific cognitive behavioral therapy; CM, contingency management.

Table I. Observed statistics and <i>p</i> -values for the initial model.									
Term	fo	а	d	q_0	р	<i>p</i> *			
(Intercept)	0.80	8.67	1.54	0.98	0.4896	0.5309			
BaseScore	15.62	0.66	1077.31	1590.96	0.0000	0.0000			
Drug	3.23	2.29	11.24	33.55	0.0005	0.0011			
HIV	1.13	9.07	1.50	1.80	0.2914	0.2966			
CBT	2.63	8.08	1.90	6.51	0.0350	0.0325			
GCBT	2.13	8.83	1.66	4.58	0.0733	0.0677			
СМ	1.02	9.59	1.43	1.44	0.3421	0.3557			

We obtain $p = P(\chi^2(d) \ge q_0)$ via the χ^2 approximation and compute $p^* = P(QF \ge f_0)$ via Davis's algorithm. CBT, cognitive behavioral therapy; GCBT, gay-specific cognitive behavioral therapy; CM, contingency management.

Table II shows the observed statistics f_0 and associated *p*-values for the reduced model. *BaseScore* and *Drug* effects are still very strong, and *Incentive* is a significant factor at 2% level. Thus, we concluded that the incentives played a contributive role here in reducing depression scores even with the existence of the *Drug* variable. This is consistent with other findings in the literature regarding CM; see [4] for example.

We further compared the reduced model with the initial model. The QF value is $f_0 = 0.91$ with $p^* = 0.53$ (associated a = 8.94, d = 4.65, $q_0 = 3.84$, and p = 0.52). We accepted the reduced model and performed pointwise diagnostics. The residuals verse fitted plots appeared to be acceptable, and the normality assumptions were reasonable.



Figure 3. Pointwise estimates and 95% confidence intervals of Incentive in the reduced model.

Table II. Observed statistics and <i>p</i> -values for the reduced model.									
Term	fo	а	d	q_0	р	p^*			
(Intercept)	2.22	8.54	1.76	5.14	0.0609	0.0560			
BaseScore	15.19	1.01	444.86	779.30	0.0000	0.0000			
Drug	3.51	2.26	12.65	38.98	0.0002	0.0005			
Incentive	3.12	8.23	2.04	8.16	0.0177	0.0172			

For comparison, we also analyzed the data by linear mixed effects (LME) models with random intercept to model heterogeneity across subjects. As seen earlier in Figure 2, the effect of *BaseScore* decreased linearly over time, whereas the effects of other variables could be assumed as constant. This was confirmed by an elaborate analysis with the consideration of possible interactions between all covariates and time. For illustration, we considered the following LME model:

$$\begin{aligned} y_{ij} &= u_i + t_j \cdot \beta_1 + BaseScore_i \cdot \beta_2 + BaseScore_i \cdot t_j \cdot \beta_3 \\ &+ HIV_i \cdot \beta_4 + Drug_{ij} \cdot \beta_5 + CBT_i \cdot \beta_6 + GCBT_i \cdot \beta_7 + CM_i \cdot \beta_8 + \epsilon_{ij}, \end{aligned}$$

where i = 1, ..., 162 indicates subjects, t_j denotes the time in week for j = 1, ..., 16, and u_i 's are the random effects explaining the heterogeneity across subjects. The parameters β s are fixed effects and do not depend on subjects. The random errors ϵ_{ij} 's are identically and independently distributed as normal. It is assumed that u_i 's are also normally distributed and independent of ϵ_{ij} . The LME model was comparable with the functional linear model (7) and led to consistent conclusions. Specifically, *BaseScore* (p < 0.0001), *BaseScore* $\cdot t$ (p < 0.0001), and *Drug* (p < .001) are very significant, *HIV* (p = .185) is not, *CBT* (p = .0104) and *GCBT* (p = .046) are different from *CBT* + CM, and *CM* (p = .291) is similar to *CBT* + CM. To compare with the reduced functional linear model (8), we also fitted another LME model

$$y_{ij} = u_i + t_j \cdot \beta_1 + BaseScore_i \cdot \beta_2 + BaseScore_i \cdot t_j \cdot \beta_3 + Drug_{ij} \cdot \beta_4 + Incentive_i \cdot \beta_5 + \epsilon_{ij}.$$

We again found that providing financial incentives for reducing methamphetamine use was very contributive to the reduction of depression scores during the study period (p = .012).

4. Simulation studies of size and power

To evaluate the performance of the proposed quasi F-test, we conducted simulation studies under similar design conditions to the aforementioned clinical trial. We focused on the power of detecting the *Incentive* effect in the functional linear model (8), which was taken as the alternative model. The null model did not include the *Incentive* term. We simulated response BDI scores as the weighted average of predicted values from the null and alternative models plus normal random errors. We used the original covariates for each subject and kept the same missing structure, that is, the simulated response was missing whenever the original BDI score was missing. For the random errors, we used two covariance structures: compound symmetric (CS) and autoregressive type 1 (AR(1)). For the CS covariance structure, the correlation coefficient between y_{ij} and y_{ik} was the same (i.e., ρ for $j \neq k$), whereas for the AR(1) case, the correlation strength depended on the distance between the two observations (i.e., $\rho^{|j-k|}$). We set the residual variance to be the average of pointwise variance estimates from model (8), which is $\sigma^2 = 1.6$. The weights varied between 0 (corresponding to the null model) and 1 (corresponding to the alternative model) with an increment of 0.1. For each weight, we generated 1000 sets of data. For each set of simulated data, we fitted both models and performed the proposed quasi *F*-tests at the significance level of 0.05. For comparison, we also fitted and tested the corresponding LME models.

Figure 4 shows the powers of the two methods for the two covariance structures with correlation set at $\rho = 0.5$ and $\rho = 0.8$. When the weight is 0, the null model is the true model and the power is the size of the test. The quasi *F*-test appears to have an actual size, and the simulated sizes range from 0.044 to 0.053. The size appears to be slightly inflated for the LME methods, ranging from 0.058 and 0.067. When the weight is relatively small (e.g., < 0.5), the LME methods have slightly higher powers than the QF method. Nevertheless, the proposed QF test has higher power than the LME method when the weight is larger than 0.8 across all cases. Both methods have higher powers for the AR(1) structure than the CS structure.

For the QF statistics, we computed both *p*-values with the χ^2 approximation approach and the exact numerical integration approach. The latter took 10 times longer than the former. The power curves were nearly identical, which suggested that the approximation is accurate and satisfactory for practical use.

5. Concluding remarks

We proposed a quasi F-test for functional linear models with functional covariates. The quasi F-statistic is a weighted average of the pointwise F-statistics, and the weights are proportional to the pointwise estimated variances. Like in the scalar case, we can use the quasi F-test to assess the significance of individual covariates in a linear model conveniently. We developed a numerical procedure and chi-squared approximation for computing p-values. The approximation is efficient in terms of both computation and accuracy; therefore, it is recommended for practical use.

We analyzed data collected from a depression and methamphetamine-dependence study, using both the QF test for functional linear models with functional covariates and the standard linear mixed effects



Figure 4. Simulated power curves of the proposed quasi F (QF) test and linear mixed effects (LME) method.

modeling method for longitudinal data. We drew similar conclusions from both methods. We further conducted simulation studies to evaluate the size and the power. The quasi F-test appears to have an accurate size and have good statistical power comparable with the commonly used linear mixed effects models.

An advantage of the quasi *F*-test is that it requires neither the specification of the within-subject covariance structure nor the functional form of coefficients. In fact, as illustrated in the application, the shape of the pointwise estimation of the coefficient functions often inspires about certain parametric form of these coefficient functions. On the other hand, a limitation of the quasi *F*-test is that the observations are needed to be taken at fixed regular time points. A sufficient number of observations are required for each time point to ensure that the estimate of covariance matrix Σ is reliable. Another potential limitation of the quasi *F*-test is its reliance on normality of the errors. It would be interesting to study the sensitivity of the quasi *F*-test to the deviation from normality.

Appendix A

Proof of Theorem 1

Let $rss_{\omega j}$ and $rss_{\Omega j}$ be the residual sums of squares of the models ω and Ω at time point t_j , respectively. From the scalar univariate linear model theory, we know $F_j = [(rss_{\omega j} - rss_{\Omega j})/(p-q)]/[rss_{\Omega j}/(n-p)]$ and $\hat{\sigma}_j^2 = rss_{\Omega j}/(n-p)$. Therefore, $(rss_{\omega j} - rss_{\Omega j})/(p-q) = F_j\hat{\sigma}_j^2$. It is evident that $rss_{\omega} = \sum_{j=1}^m rss_{\omega j}$ and $rss_{\Omega} = \sum_{j=1}^m rss_{\Omega j}$. So, $(rss_{\omega} - rss_{\Omega})/(p-q) = \sum_{j=1}^m F_j\hat{\sigma}_j^2$ and the QF statistic in Equation (3) is $QF = \sum_{j=1}^m F_j\hat{\sigma}_j^2/(rss_{\Omega}/(n-p))$. This completes the proof. \Box

Acknowledgements

The authors gratefully acknowledge the support from NIDA grants 1R01DA11031 and 1P50DA18185. The authors also thank an associate editor and two referees for their comments.

References

- 1. Hand DJ, Crowder MJ. Practical Longitudinal Data Analysis. Chapman and Hall: London, 1996.
- 2. Diggle PJ, Heagerty P, Liang KY, Zeger SL. Analysis of Longitudinal Data, 2nd ed. University Press: Oxford, 2002.
- Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. Longitudinal Data Analysis, (Handbooks of Modern Statistical Methods). Chapman & Hall/CRC: New York, 2009.
- Shoptaw S, Rotheram-Fuller E, Yang X, Frosch D, Nahom D, Jarvik ME, Rawson RA, Ling W. Smoking cessation in methadone maintenance. *Addiction* 2002; 97:1317–1328.
- Peck JA, Reback CJ, Yang X, Rotheram-Fuller E, Shoptaw S. Sustained reductions in drug use and depression symptoms from treatment for drug abuse in methamphetamine-dependent gay and bisexual men. *Journal of Urban Health: Bulletin* of the New York Academy of Medicine 2005; 82(Supplement 1):i100–i108.
- 6. Ramsay JO, Silverman BW. Functional Data Analysis, 2nd ed. Springer: New York, 2005.
- 7. Faraway JJ. Regression analysis for a functional response. Technometrics 1997; 39:254-261.
- 8. Fan J, Lin S-K. Tests of significance when data are curves. *Journal of the American Statistical Association* 1998; **93**:1007–1021.
- 9. Shen Q, Faraway J. An F test for linear models with functional responses. Statistica Sinica 2004; 14:1239–1257.
- 10. Shen Q, Xu H. Diagnostics for linear models with functional responses. Technometrics 2007; 49:26–33.
- 11. Yang X, Shen Q, Xu H, Shoptaw S. Functional regression analysis using an F test for longitudinal data with large numbers of repeated measures. *Statistics in Medicine* 2007; **26**:1552–1566.
- 12. Yang X, Nie K. Hypothesis testing in functional linear regression models with Neyman's truncation and wavelet thresholding for longitudinal data. *Statistics in Medicine* 2008; **27**:845–863.
- 13. Zhang J-T, Chen J. Statistical inferences for functional data. Annals of Statistics 2007; 35:1052–1079.
- 14. Brumback BA, Rice JA. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association* 1998; **93**:961–976.
- Hoover DR, Rice JA, Wu CO, Yang L-P. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 1998; 85:809–822.
- 16. Hastie T, Tibshirani R. Varying-coefficient models (with discussions). *Journal of the Royal Statistical Society, Series B* 1993; **55**:757–796.
- 17. Wu CO, Yu KF. Nonparametric varying-coefficient models for the analysis of longitudinal data. *International Statistical Review* 2002; **70**:373–393.
- 18. Fan J, Zhang W. Statistical methods with varying coefficient models. *Statistics and its Interface* 2008; **1**:179–195.
- 19. Satterthwaite FE. An approximate distribution of estimates of variance components. *Biometrics Bulletin* 1946; 2:110–114.
- 20. Imhof JP. Computing the distribution of quadratic forms in normal variables. *Biometrika* 1961; **48**:419–426.



- 21. Davis RB. Algorithm AS 155: the distribution of a linear combination of χ^2 random variables. *Journal of the Royal Statistical Society, Series C* 1980; **29**:323–333.
- 22. Pearson ES. Note on an approximation to the distribution of non-central χ^2 . *Biometrika* 1959; **46**:364.
- 23. Zhang J-T. Approximate and asymptotic distributions of chi-squared-type mixtures with applications. *Journal of the American Statistical Association* 2005; **100**:273–285.
- 24. Seber GAF, Lee AJ. Linear Regression Analysis, 2nd edn. Wiley: New York, 2003.
- 25. Abraham B, Ledolter J. Introduction to Regression Modeling. Duxbury Press: Belmont, CA, 2006.
- 26. Muller KE, Stewart PW. Linear Model Theory: Univariate, Multivariate, and Mixed Models. Wiley: New York, 2006.
- 27. Box GEP. Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics* 1954; **25**:290–302.