Tests for Error Correlation in the Functional Linear Model

Robertas GABRYS, Lajos HORVÁTH, and Piotr KOKOSZKA

The paper proposes two inferential tests for error correlation in the functional linear model, which complement the available graphical goodness-of-fit checks. To construct them, finite dimensional residuals are computed in two different ways, and then their autocorrelations are suitably defined. From these autocorrelation matrices, two quadratic forms are constructed whose limiting distribution are chi-squared with known numbers of degrees of freedom (different for the two forms). The asymptotic approximations are suitable for moderate sample sizes. The test statistics can be relatively easily computed using the `R` package `fda`, or similar MATLAB software. Application of the tests is illustrated on magnetometer and financial data. The asymptotic theory emphasizes the differences between the standard vector linear regression and the functional linear regression. To understand the behavior of the residuals obtained from the functional linear model, the interplay of three types of approximation errors must be considered, whose sources are: projection on a finite dimensional subspace, estimation of the optimal subspace, and estimation of the regression kernel.

KEY WORDS: Correlated errors; Functional regression; Principal components.

1. INTRODUCTION

The last decade has seen the emergence of the functional data analysis (FDA) as a useful area of statistics which provides convenient and informative tools for the analysis of data objects of large dimension. The influential book of Ramsay and Silverman (2005) provides compelling examples of the usefulness of this approach. Functional data arise in many contexts. This paper is motivated by our work with data obtained from very precise measurements at fine temporal grids which arise in engineering, physical sciences, and finance. At the other end of the spectrum are sparse data measured with error which are transformed into curves via procedures that involve smoothing. Such data arise, for example, in longitudinal studies on human subjects or in biology, and wherever frequent, precise measurements are not feasible. Our methodology and theory are applicable to such data after they have been appropriately transformed into functional curves. Many such procedures are now available.

Like its classical counterpart, the functional linear model stands out as a particularly useful tool, and has consequently been thoroughly studied and extensively applied; see Cuevas, Febrero, and Fraiman (2002), Malfait and Ramsay (2003), Cardot et al. (2003), Chiou, Müller, and Wang (2004), Müller and Stadtmüller (2005), Yao, Müller, and Wang (2005a, 2005b), Cai and Hall (2006), Chiou and Müller (2007), Li and Hsing (2007), Reiss and Ogden (2007, 2009, 2010), among many others.

For any statistical model, it is important to evaluate its suitability for particular data. In the context of the multivariate linear regression, well-established approaches exist, but for the functional linear model, only the paper of Chiou and Müller (2007) addresses the diagnostics in any depth. These authors emphasize the role of the functional residuals $\hat{\varepsilon}_i(t) = \hat{Y}_i(t) - Y_i(t)$, where the $Y_i(t)$ are the response curves, and the $\hat{Y}_i(t)$ are the fitted curves, and propose a number of graphical tools, akin to the usual residual plots, which offer a fast and convenient way of assessing the goodness of fit. They also propose a test statistic based on Cook's distance, Cook (1977) or Cook and Weisberg (1982), whose null distribution can be computed by randomizing a binning scheme.

We propose two goodness-of-fit tests aimed at detecting serial correlation in the error functions $\varepsilon_n(t)$ in the fully functional model

$$Y_n(t) = \int \psi(t,s) X_n(s)\, ds + \varepsilon_n(t), \qquad n = 1, 2, \ldots, N. \quad (1.1)$$

The assumption of iid ε_n underlies all inferential procedures for model (1.1) proposed to date. As in the multivariate regression, error correlation affects various variance estimates, and, consequently, confidence regions and distributions of test statistics. In particular, prediction based on LS estimation is no longer optimal. In the context of scalar data, these facts are well known and go back at least to Cochrane and Orcutt (1949). If functional error correlation is detected, currently available inferential procedures cannot be used. At this point, no inferential procedures for the functional linear model with correlated errors are available, and it is hoped that this paper will motivate research in this direction. For scalar data, the relevant research is very extensive, so we mention only the influential papers of Sacks and Ylvisaker (1966) and Rao and Griliches (1969), and refer to textbook treatments in chapters 9 and 10 of Seber and Lee (2003), chapter 8 of Hamilton (1989), and section 13.5 of Bowerman and O'Connell (1990). The general idea is that when dependence in errors is detected, it must be modeled, and inference must be suitably adjusted.

The methodology of Chiou and Müller (2007) was not designed to detect error correlation, and can leave it undetected. Figure 1 shows diagnostic plots of Chiou and Müller (2007) obtained for synthetic data that follow a functional linear model with highly correlated errors. These plots exhibit almost ideal football shapes. It is equally easy to construct examples in which our methodology fails to detect departures from model

Robertas Gabrys is a Doctoral Candidate in Statistics, Department of Mathematics and Statistics, Utah State University, Logan, UT 84322 (E-mail: *gabrys.robertas@gmail.com*). Lajos Horváth is Professor, Department of Mathematics, University of Utah, Salt Lake City, UT 84112. Piotr Kokoszka is Professor, Department of Mathematics and Statistics, Utah State University, Logan, UT 84322. The research was partially supported by NSF grants DMS-0804165 and DMS-0931948, at Utah State University, and DMS-0905400, at the University of Utah. Three referees provided plentiful comments which have led to substantive improvements. We thank the Associate Editor and the Editor for providing a timely and focused evaluation.

Journal of the American Statistical Association
September 2010, Vol. 105, No. 491, Theory and Methods
DOI: 10.1198/jasa.2010.tm09794

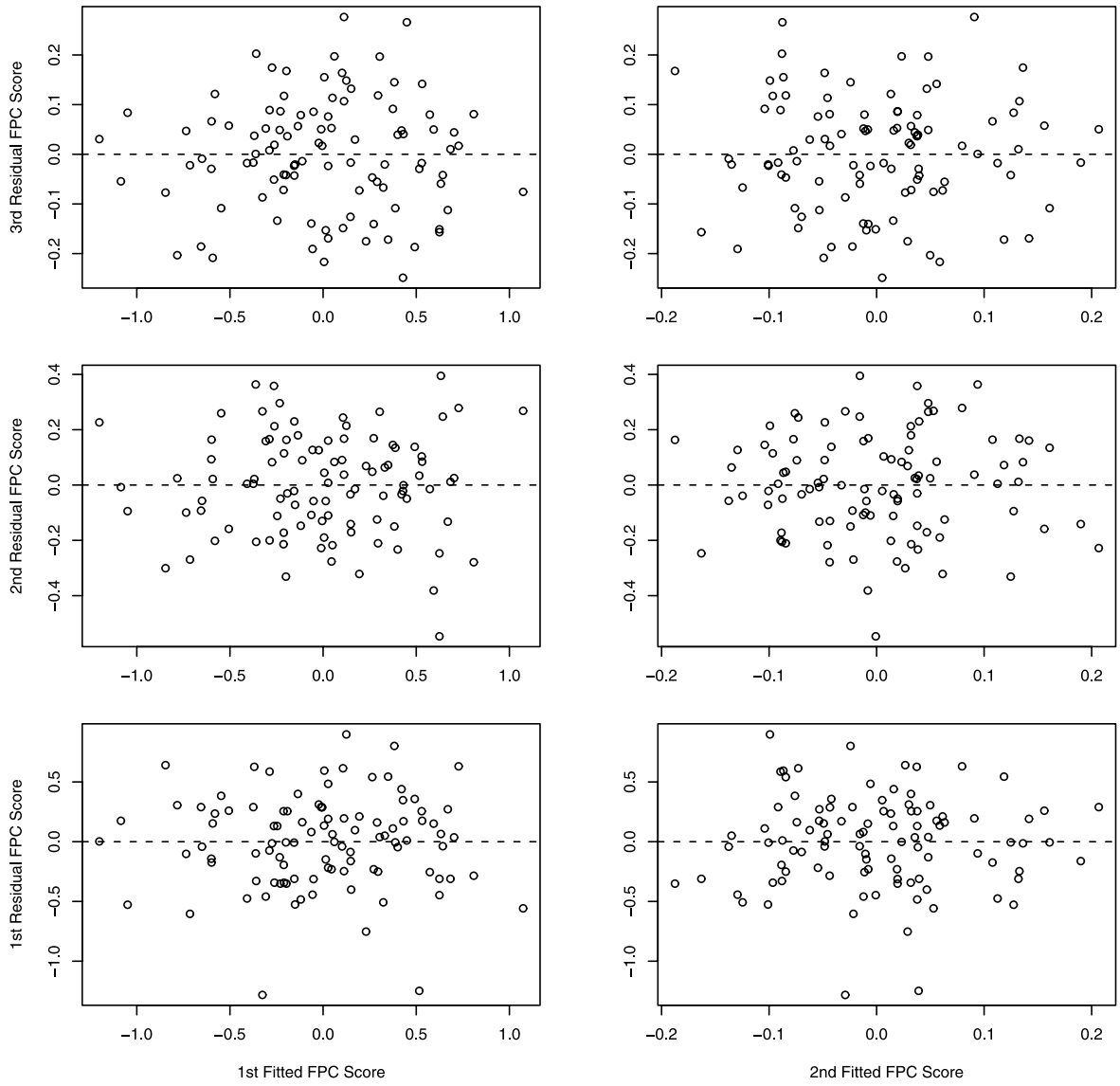

Figure 1. Diagnostic plots of Chiou and Müller (2007) for a synthetic dataset simulated according to model (1.1) in which the errors ε_n follow the functional autoregressive model of Bosq (2000).

(1.1), but the graphs of Chiou and Müller (2007) immediately show it. The simplest such example is given by $Y_n(t) = X_n^2(t) + \varepsilon_n(t)$ with iid ε_n. Thus, the methods we propose are complimentary tools designed to test the validity of specification (1.1) with iid errors against the alternative of correlation in the errors.

Despite a complex asymptotic theory, the null distribution of both test statistics we propose is asymptotically chi-squared, which turns out to be a good approximation in finite samples. The test statistics are relatively easy to compute; an R code is available upon request. They can be viewed as nontrivial refinements of the ideas of Durbin and Watson (1950, 1951, 1971); see also Chatfield (1998) and section 10.4.4 of Seber and Lee (2003), who introduced tests for serial correlation in the standard linear regression. Their statistics are functions of sample autocorrelations of the residuals, but their asymptotic distributions depend on the distribution of the regressors, and so various additional steps and rough approximations are required; see Thiel and Nagar (1961) and Thiel (1965), among others. To overcome these difficulties, Schmoyer (1994) proposed permutation tests based on quadratic forms of the residuals. We appropriately define residual autocorrelations, and their quadratic forms (not the quadratic forms of the residuals as in Schmoyer 1994), in such a way that the asymptotic distribution is the standard chi-squared distribution.

The complexity of the requisite asymptotic theory is due to the fact that in order to construct a computable test statistic, finite dimensional objects reflecting the relevant properties of the infinite dimensional unobservable errors $\varepsilon_n(t)$ must be constructed. In the standard regression setting, the explanatory variables live in a finite dimensional Euclidean space with a fixed (standard) basis, and the residuals reflect the effect of parameter estimation. In the functional setting, before any estimation can be undertaken, the dimension of the data must be reduced, typically by projecting on an "optimal" finite dimensional subspace. This projection operation introduces an error. Next, the "optimal subspace" must be estimated, and this introduces another error. Finally, estimation of the kernel $\psi(\cdot, \cdot)$ introduces

still another error. Our asymptotic approach focuses on the impact of these errors. We do not consider the dimensions of the optimal projection spaces growing to infinity with the sample size. Such an asymptotic analysis is much more complex; in a simpler setting it was developed by Panaretos, Kraus, and Maddocks (2010).

The two methods proposed in this paper start with two ways of defining the residuals. Method I uses projections of all curves on the functional principal components of the regressors, and so is closer to the standard regression in that one common basis is used. This approach is also useful for testing the stability of model (1.1) against a change point alternative; see Horváth, Kokoszka, and Maslova (2009). Method II uses two bases: the eigenfunctions of the covariance operators of the regressors and of the responses.

The remainder of the paper is organized as follows. Section 2 introduces the assumptions and the notation. Section 3 develops the setting for the least squares estimation needed to define the residuals used in Method I. After these preliminaries, both tests are described in Section 4, with the asymptotic theory presented in Section 5. The finite sample performance is evaluated in Section 6 through a simulation study, and further examined in Section 7 by applying both methods to magnetometer and financial data. All proofs are collected in the *extended version* available at *www.stat.usu.edu/~piotr/gfExt.pdf*.

2. PRELIMINARIES

We denote by L^2 the space of square integrable functions on the unit interval, and by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ the usual inner product and the norm it generates.

The usual conditions imposed on model (1.1) are collected in the following assumption.

Assumption 1. The errors ε_n are independent identically distributed mean zero elements of L^2 satisfying $E\|\varepsilon_n\|^4 < \infty$. The covariates X_n are independent identically distributed mean zero elements of L^2 satisfying $E\|X_n\|^4 < \infty$. The sequences $\{X_n\}$ and $\{\varepsilon_n\}$ are independent.

For data collected sequentially over time, the regressors X_n need not be independent. We formalize the notion of dependence in functional observations using the notion of L^4-m-approximability advocated in other contexts by Hörmann (2008), Berkes, Hörmann, and Schauer (2009), Aue et al. (2009), and used for functional data by Hörmann and Kokoszka (2010) and Aue et al. (2010). We now list the assumptions we need to establish the asymptotic theory. For ease of reference, we repeat some conditions contained in Assumption 1; the weak dependence of the $\{X_n\}$ is quantified in Conditions (A2) and (A5). Assumption 1 will be needed to state intermediate results.

(A1) The ε_n are independent, identically distributed with $E\varepsilon_n = 0$ and $E\|\varepsilon_n\|^4 < \infty$.

(A2) Each X_n admits the representation

$$X_n = g(\alpha_n, \alpha_{n-1}, \ldots),$$

in which the α_k are independent, identically distributed elements of a measurable space S, and $g : S^\infty \to L^2$ is a measurable function.

(A3) The sequences $\{\varepsilon_n\}$ and $\{\alpha_n\}$ are independent.

(A4) $EX_n = 0$, $E\|X_n\|^4 < \infty$.

(A5) There are $c_0 > 0$ and $\kappa > 2$ such that

$$\left(E\left\|X_n - X_n^{(k)}\right\|^4\right)^{1/4} \le c_0 k^{-\kappa},$$

where

$$X_n^{(k)} = g\left(\alpha_n, \alpha_{n-1}, \ldots, \alpha_{n-k+1}, \alpha_{n-k}^{(k)}, \alpha_{n-k-1}^{(k)}, \ldots\right),$$

and where the $\alpha_\ell^{(k)}$ are independent copies of α_0.

Condition (A2) means that the sequence $\{X_n\}$ admits a causal representation known as a Bernoulli shift. It follows from (A2) that $\{X_n\}$ is stationary and ergodic. The structure of the function $g(\cdot)$ is not important; it can be a linear or a highly nonlinear function. What matters is that according to (A5), $\{X_n\}$ is weakly dependent, as it can be approximated with sequences of k-dependent variables, and the approximation improves as k increases. Several examples of functional sequences satisfying (A2), (A4), and (A5) can be found in Hörmann and Kokoszka (2010) and Aue et al. (2010). They include functional linear, bilinear, and conditionally heteroscedastic processes.

We denote by C the covariance operator of the X_i defined by $C(x) = E[\langle X, x\rangle X]$, $x \in L^2$, where X has the same distribution as the X_i. By λ_k and v_k, we denote, correspondingly, the eigenvalues and the eigenfunctions of C. The corresponding objects for the Y_i are denoted Γ, γ_k, u_k, so that

$$C(v_k) = \lambda_k v_k, \qquad X_n = \sum_{i=1}^{\infty} \xi_{ni} v_i, \qquad \xi_{ni} = \langle v_i, X_n \rangle;$$

$$\Gamma(u_k) = \gamma_k u_k, \qquad Y_n = \sum_{j=1}^{\infty} \zeta_{nj} u_j, \qquad \zeta_{nj} = \langle u_j, Y_n \rangle.$$

In practice, we must replace the population eigenfunctions and eigenvalues by their empirical counterparts $\hat{\lambda}_k, \hat{v}_k, \hat{\gamma}_k, \hat{u}_k$ defined as the eigenelements of the empirical covariance operators [we assume $EX_n(t) = 0$]

$$\hat{C}(x) = N^{-1} \sum_{n=1}^{N} \langle X_n, x \rangle X_n, \qquad x \in L^2,$$

and analogously defined $\hat{\Gamma}$. The empirical scores are also denoted with the "hat," that is, by $\hat{\xi}_{ni}$ and $\hat{\zeta}_{nj}$. We often refer to the v_i, u_j as the functional principal components (FPC's), and to the $\hat{v}_i, \hat{u}_j$ as the empirical functional principal components (EFPC's).

To state the alternative, we must impose dependence conditions on the ε_n. We use the same conditions that we imposed on the X_n, because then the asymptotic arguments under H_A can use the results derived for the X_n under H_0. Specifically, we introduce the following assumptions:

(B1) $E\varepsilon_n = 0$ and $E\|\varepsilon_n\|^4 < \infty$.

(B2) Each ε_n admits the representation

$$\varepsilon_n = h(u_n, u_{n-1}, \ldots),$$

in which the u_k are independent, identically distributed elements of a measurable space S, and $h : S^\infty \to L^2$ is a measurable function.

(B3) The sequences $\{u_n\}$ and $\{\alpha_n\}$ are independent.

(B4) There are $c_0 > 0$ and $\kappa > 2$ such that

$$\left(E\left\|\varepsilon_n - \varepsilon_n^{(k)}\right\|^4\right)^{1/4} \le c_0 k^{-\kappa},$$

where

$$\varepsilon_n^{(k)} = h\left(u_n, u_{n-1}, \ldots, u_{n-k+1}, u_{n-k}^{(k)}, u_{n-k-1}^{(k)}, \ldots\right),$$

and where the $u_\ell^{(k)}$ are independent copies of u_0.

The tests proposed in Section 4 detect dependence which manifests itself in a correlation between ε_n and ε_{n+h} for at least one h. Following Bosq (2000), we say that ε_n and ε_{n+h} are uncorrelated if $E[\langle \varepsilon_n, x\rangle\langle \varepsilon_{n+h}, y\rangle] = 0$ for all $x, y \in L^2$. If $\{e_j\}$ is any orthonormal basis in L^2, this is equivalent to $E[\langle \varepsilon_n, e_i\rangle\langle \varepsilon_{n+h}, e_j\rangle] = 0$ for all i, j. The two methods introduced in Section 4 detect the alternatives with $e_i = v_i$ (Method I) and $e_i = u_i$ (Method II). These methods test for correlation up to lag H, and use the FPC v_i, $i \le p$, and u_i, $i \le q$.

With this background, we can state the null and alternative hypotheses as follows.

H_0: Model (1.1) holds together with Assumptions (A1)–(A5).

The key assumption is (A1), that is, the independence of the ε_n.

$H_{A,I}$: Model (1.1) holds together with Assumptions (A2), (A4), (A5), (B1)–(B4), and $E[\langle \varepsilon_0, v_i\rangle\langle \varepsilon_h, v_j\rangle] \ne 0$ for some $1 \le h \le H$ and $1 \le i, j \le p$.

$H_{A,II}$: Model (1.1) holds together with Assumptions (A2), (A4), (A5), (B1)–(B4), and $E[\langle \varepsilon_0, u_i\rangle\langle \varepsilon_h, u_j\rangle] \ne 0$ for some $1 \le h \le H$ and $1 \le i, j \le q$.

Note that the u_i are well defined under the alternative, because (A2), (A4), (A5), and (B1)–(B4) imply that the Y_n form a stationary sequence.

In the proofs, we will often use the following result established in Hörmann and Kokoszka (2010) and Aue et al. (2010). In Theorem 1, and in the following, we set

$$\hat{c}_j = \operatorname{sign}(\langle v_j, \hat{v}_j\rangle).$$

Theorem 1. Suppose Assumptions (A2), (A4), and (A5) hold, and

$$\lambda_1 > \lambda_2 > \cdots > \lambda_p > \lambda_{p+1}. \tag{2.1}$$

Then, for each $1 \le j \le p$,

$$\begin{aligned} &\limsup_{N\to\infty} NE[\|\hat{c}_j\hat{v}_j - v_j\|^2] < \infty, \\ &\limsup_{N\to\infty} NE[|\lambda_j - \hat{\lambda}_j|^2] < \infty. \end{aligned} \tag{2.2}$$

3. LEAST SQUARES ESTIMATION

In this section we show how model (1.1) can be cast into a standard estimable form. The idea is different from the usual approaches, for example, in Ramsay and Silverman (2005) and Yao, Müller, and Wang (2005b), so a detailed exposition is necessary. The goal is to obtain clearly defined residuals which can be used to construct a goodness-of-fit test. This section carefully explains the three steps involved in the construction of the residuals in the setting of model (1.1). The idea is that the curves are represented by their coordinates with respect to the FPC's of the X_n, for example, $Y_{nk} = \langle Y_n, v_k\rangle$ is the projection of the nth response onto the kth largest FPC. A formal linear model for these coordinates is constructed and estimated by least squares. This formal model does not, however, satisfy the usual assumptions due to the effect of the projection of infinite dimensional curves on a finite dimensional subspace, and so its asymptotic analysis is delicate.

Since the v_k form a basis in $L^2([0,1])$, the products $v_i(t)v_j(s)$ form a basis in $L^2([0,1]\times[0,1])$. Thus, if $\psi(\cdot,\cdot)$ is a Hilbert–Schmidt kernel, then

$$\psi(t,s) = \sum_{i,j=1}^{\infty} \psi_{ij} v_i(t) v_j(s), \tag{3.1}$$

where $\psi_{ij} = \int \psi(t,s) v_i(t) v_j(s)\, dt\, ds$. Therefore,

$$\int \psi(t,s) X_n(s)\, ds = \sum_{i,j=1}^{\infty} \psi_{ij} v_i(t) \langle X_n, v_j\rangle.$$

Hence, for any $1 \le k \le p$, we have

$$Y_{nk} = \sum_{j=1}^{p} \psi_{kj}\xi_{nj} + e_{nk} + \eta_{nk}, \tag{3.2}$$

where

$$Y_{nk} = \langle Y_n, v_k\rangle, \qquad \xi_{nj} = \langle X_n, v_j\rangle, \qquad e_{nk} = \langle \varepsilon_n, v_k\rangle,$$

and where

$$\eta_{nk} = \sum_{j=p+1}^{\infty} \psi_{kj}\langle X_n, v_j\rangle.$$

We combine the errors e_{nk} and η_{nk} by setting

$$\delta_{nk} = e_{nk} + \eta_{nk}.$$

Note that the δ_{nk} are no longer iid.

Setting

$$\begin{aligned} &\mathbf{X}_n = [\xi_{n1}, \ldots, \xi_{np}]^T, \qquad \mathbf{Y}_n = [Y_{n1}, \ldots, Y_{np}]^T, \\ &\boldsymbol{\delta}_n = [\delta_{n1}, \ldots, \delta_{np}]^T, \\ &\boldsymbol{\psi} = [\psi_{11}, \ldots, \psi_{1p}, \psi_{21}, \ldots, \psi_{2p} \ldots, \psi_{p1}, \ldots, \psi_{pp}]^T, \end{aligned}$$

we rewrite (3.2) as

$$\mathbf{Y}_n = \mathbf{Z}_n\boldsymbol{\psi} + \boldsymbol{\delta}_n, \qquad n = 1, 2, \ldots, N,$$

where each $\mathbf{Z}_n$ is a $p \times p^2$ matrix

$$\mathbf{Z}_n = \begin{bmatrix} \mathbf{X}_n^T & \mathbf{0}_p^T & \cdots & \mathbf{0}_p^T \\ \mathbf{0}_p^T & \mathbf{X}_n^T & \cdots & \mathbf{0}_p^T \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_p^T & \mathbf{0}_p^T & \cdots & \mathbf{X}_n^T \end{bmatrix}$$

with $\mathbf{0}_p = [0, \ldots, 0]^T$.

Finally, defining the $Np \times 1$ vectors $\mathbf{Y}$ and $\boldsymbol{\delta}$ and the $Np \times p^2$ matrix $\mathbf{Z}$ by

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix}, \qquad \boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_N \end{bmatrix}, \qquad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_N \end{bmatrix},$$

we obtain the following linear model

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\psi} + \boldsymbol{\delta}. \tag{3.3}$$

Note that (3.3) is not a standard linear model. Firstly, the design matrix $\mathbf{Z}$ is random. Secondly, $\mathbf{Z}$ and $\boldsymbol{\delta}$ are not independent. The error term $\boldsymbol{\delta}$ in (3.3) consists of two parts: the projections of the ε_n, and the remainder of an infinite sum. Thus, while (3.3) looks like the standard linear model, the existing asymptotic results do not apply to it, and a new asymptotic analysis involving the interplay of the various approximation errors is needed. Representation (3.3) leads to the formal "least squares estimator" for $\boldsymbol{\psi}$ is

$$\hat{\boldsymbol{\psi}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y} = \boldsymbol{\psi} + (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\boldsymbol{\delta}, \tag{3.4}$$

which cannot be computed because the v_k must be replaced by the $\hat{v}_k$.

Now we turn to the effect of replacing the v_k by the $\hat{v}_k$. Projecting onto the $\hat{v}_k$, we are "estimating" the *random* vector

$$\tilde{\boldsymbol{\psi}} = [\hat{c}_1\psi_{11}\hat{c}_1, \ldots, \hat{c}_1\psi_{1p}\hat{c}_p, \ldots, \hat{c}_p\psi_{p1}\hat{c}_1, \ldots, \hat{c}_p\psi_{pp}\hat{c}_p]^T \tag{3.5}$$

with the "estimator"

$$\tilde{\boldsymbol{\psi}}^{\wedge} = (\hat{\mathbf{Z}}^T\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}^T\hat{\mathbf{Y}}$$

obtained by replacing the v_k by the $\hat{v}_k$ in (3.4). It will be convenient to associate this vector of dimension p^2 with the $p \times p$ matrix

$$\tilde{\boldsymbol{\Psi}}_p^{\wedge} = \begin{bmatrix} \tilde{\psi}_{11}^{\wedge} & \tilde{\psi}_{12}^{\wedge} & \cdots & \tilde{\psi}_{1p}^{\wedge} \\ \tilde{\psi}_{21}^{\wedge} & \tilde{\psi}_{22}^{\wedge} & \cdots & \tilde{\psi}_{2p}^{\wedge} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\psi}_{p1}^{\wedge} & \tilde{\psi}_{p2}^{\wedge} & \cdots & \tilde{\psi}_{pp}^{\wedge} \end{bmatrix}. \tag{3.6}$$

It can be shown that if the regularity conditions of Hall and Hosseini-Nasab (2006) hold, then

$$N^{1/2}(\tilde{\boldsymbol{\psi}}^{\wedge} - \tilde{\boldsymbol{\psi}}) = [\widehat{\mathbf{C}} \otimes \widehat{\mathbf{C}}]N^{1/2}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) + \mathbf{Q}^{-1}(R_{N1} + R_{N2}) + o_P(1), \tag{3.7}$$

where

$$\widehat{\mathbf{C}} = \begin{bmatrix} \hat{c}_1 & 0 & \cdots & 0 \\ 0 & \hat{c}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \hat{c}_p \end{bmatrix}, \qquad \mathbf{Q} = \mathbf{I}_p \otimes \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}, \tag{3.8}$$

and where $\otimes$ denotes the Kronecker product of two matrices. The terms R_{N1} and R_{N2} are linear functionals of $N^{-1/2} \times \sum_{n=1}^{N} X_n(t)$ and $N^{-1/2}\sum_{n=1}^{N}\{X_n(t)X_n(s) - E[X_n(t)X_n(s)]\}$. The limits of $N^{1/2}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})$, R_{N1}, and R_{N2} are thus jointly Gaussian, but the asymptotic normality of $\tilde{\boldsymbol{\psi}}^{\wedge} - \tilde{\boldsymbol{\psi}}$ does not follow due to the random signs $\hat{c}_j$. It does however follow from (3.7) that $N^{1/2}(\tilde{\boldsymbol{\psi}}^{\wedge} - \tilde{\boldsymbol{\psi}}) = O_P(1)$, and this relation does not require the regularity assumptions of Hall and Hosseini-Nasab (2006). The rate $N^{1/2}$ is optimal, that is, if $a_N/N^{1/2} \to \infty$, then $a_N(\tilde{\boldsymbol{\psi}}^{\wedge} - \tilde{\boldsymbol{\psi}}) \xrightarrow{P} \infty$. This is exactly the result that will be used in the following, and we state it here as Proposition 1. We need the following additional assumption.

Assumption 2. The coefficients ψ_{ij} of the kernel $\psi(\cdot,\cdot)$ satisfy $\sum_{i,j=1}^{\infty}|\psi_{ij}| < \infty$.

Proposition 1. If Assumptions (A1)–(A5) and 2 hold, then $\tilde{\boldsymbol{\psi}}^{\wedge} - \tilde{\boldsymbol{\psi}} = O_P(N^{-1/2})$.

The proof of Proposition 1 is fairly technical and is developed in Aue et al. (2010).

Relation (3.7) shows that replacing the v_k by the $\hat{v}_k$ changes the asymptotic distribution. While the limiting distribution of $\boldsymbol{\psi}^{\wedge}$ is complex and cannot be used directly, this estimator itself can be used to construct a feasible goodness-of-fit test.

4. TESTING THE INDEPENDENCE OF MODEL ERRORS

We propose two test statistics, (4.5) and (4.8), which can be used to test the assumption that the errors ε_n in (1.1) are iid functions in L^2. These statistics arise from two different ways of defining finite dimensional vectors of residuals. Method I builds on the ideas presented in Section 3, the residuals are derived using the estimator $\tilde{\boldsymbol{\psi}}^{\wedge}$ obtained by projecting both the Y_n and the X_n on the $\hat{v}_i$, the functional principal components of the regressors. Method II uses two projections. As before, the X_n are projected on the $\hat{v}_i$, but the Y_n are projected on the $\hat{u}_i$. Thus, as in Yao, Müller, and Wang (2005b), we approximate $\psi(\cdot,\cdot)$ by

$$\hat{\psi}_{pq}(t,s) = \sum_{j=1}^{q}\sum_{i=1}^{p}\hat{\lambda}_i^{-1}\hat{\sigma}_{ij}\hat{u}_j(t)\hat{v}_i(s), \qquad \hat{\sigma}_{ij} = N^{-1}\sum_{n=1}^{N}\langle X_n, \hat{v}_i\rangle\langle Y_n, \hat{u}_j\rangle. \tag{4.1}$$

Method I emphasizes the role of the regressors X_n, and is, in a very loose sense, analogous to the plot of the residuals against the independent variable in a straight line regression. Method II emphasizes the role of the responses, and is somewhat analogous to the plot of the residuals against the fitted values. Both statistics have the form $\sum_{h=1}^{H}\hat{\mathbf{r}}_h^T\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{r}}_h$, where $\hat{\mathbf{r}}_h$ are vectorized covariance matrices of appropriately constructed residuals, and $\hat{\boldsymbol{\Sigma}}$ is a suitably constructed matrix which approximates the covariance matrix of the the $\hat{\mathbf{r}}_h$, which are asymptotically iid. As in all procedures of this type, the p-values are computed for a range of values of H, typically $H \le 5$ or $H \le 10$. The main difficulty, and a central contribution of this paper, is in deriving explicit formulas for the $\hat{\mathbf{r}}_h$ and $\hat{\boldsymbol{\Sigma}}$ and showing that the test statistics converge to the χ^2 distribution despite a very complex structure of the residuals in the fully functional linear model.

Method I. Recall the definition of the matrix $\tilde{\boldsymbol{\Psi}}_p^{\wedge}$ (3.6) whose (i,j) entry approximates $\hat{c}_i\psi_{ij}\hat{c}_j$, and define also $p \times 1$ vectors

$$\hat{\mathbf{Y}}_n = [\hat{Y}_{n1}, \hat{Y}_{n2}, \ldots, \hat{Y}_{np}]^T, \qquad \hat{Y}_{nk} = \langle Y_n, \hat{v}_k\rangle;$$
$$\hat{\mathbf{X}}_n = [\hat{\xi}_{n1}, \hat{\xi}_{n2}, \ldots, \hat{\xi}_{np}]^T, \qquad \hat{\xi}_{nk} = \langle X_n, \hat{v}_k\rangle.$$

The fitted vectors are then

$$\tilde{\mathbf{Y}}_n^{\wedge} = \tilde{\boldsymbol{\Psi}}_p^{\wedge}\hat{\mathbf{X}}_n, \qquad n = 1, 2, \ldots, N, \tag{4.2}$$

and the residuals are $\mathbf{R}_n = \hat{\mathbf{Y}}_n - \tilde{\mathbf{Y}}_n^{\wedge}$. For $0 \le h < N$, define the sample autocovariance matrices of these residuals as

$$\mathbf{V}_h = N^{-1} \sum_{n=1}^{N-h} \mathbf{R}_n \mathbf{R}_{n+h}^T. \tag{4.3}$$

Finally, by $\operatorname{vec}(\mathbf{V}_h)$ denote the column vectors of dimension p^2 obtained by stacking the columns of the matrices $\mathbf{V}_h$ on top of each other starting from the left. Next, define

$$e_{nk}^{\wedge} = \langle Y_n, \hat{v}_k \rangle - \sum_{j=1}^{p} \tilde{\psi}_{kj}^{\wedge} \langle X_n, \hat{v}_j \rangle,$$

$$\widehat{\mathbf{M}}_0 = \left[\frac{1}{N} \sum_{n=1}^{N} e_{nk}^{\wedge} e_{nk'}^{\wedge},\ 1 \le k, k' \le p \right],$$

and

$$\widehat{\mathbf{M}} = \widehat{\mathbf{M}}_0 \otimes \widehat{\mathbf{M}}_0. \tag{4.4}$$

With this notation in place, we can define the test statistic

$$Q_N^{\wedge} = N \sum_{h=1}^{H} [\operatorname{vec}(\mathbf{V}_h)]^T \widehat{\mathbf{M}}^{-1} \operatorname{vec}(\mathbf{V}_h). \tag{4.5}$$

Properties of the Kronecker product, $\otimes$, give simplified formulae for $Q_N^{\wedge}$. Since $\widehat{\mathbf{M}}^{-1} = \widehat{\mathbf{M}}_0^{-1} \otimes \widehat{\mathbf{M}}_0^{-1}$, see Horn and Johnson (1991, p. 244), by problem 25 on p. 252 of Horn and Johnson (1991), we have

$$Q_N^{\wedge} = N \sum_{h=1}^{H} \operatorname{tr}[\widehat{\mathbf{M}}_0^{-1} \mathbf{V}_h^T \widehat{\mathbf{M}}_0^{-1} \mathbf{V}_h].$$

Denoting by $\hat{m}_{f,h}(i,j)$ and $\hat{m}_{b,h}(i,j)$ the (i,j) entries, respectively, of $\widehat{\mathbf{M}}^{-1}\mathbf{V}_h$ and $\mathbf{V}_h\widehat{\mathbf{M}}^{-1}$, we can write according to the definition of the trace

$$Q_N^{\wedge} = N \sum_{h=1}^{H} \sum_{i,j=1}^{p} \hat{m}_{f,h}(i,j) \hat{m}_{b,h}(i,j).$$

The null hypothesis is rejected if $Q_N^{\wedge}$ exceeds an upper quantile of the chi-squared distribution with p^2H degrees of freedom; see Theorem 2.

Method II. Equation (1.1) can be rewritten as

$$\sum_{j=1}^{\infty} \zeta_{nj} u_j = \sum_{i=1}^{\infty} \xi_{ni} \Psi(v_i) + \varepsilon_n, \tag{4.6}$$

where Ψ is the Hilbert–Schmidt operator with kernel $\psi(\cdot,\cdot)$. To define the residuals, we replace the infinite sums in (4.6) by finite sums, the unobservable u_j, v_i with the $\hat{u}_j, \hat{v}_i$, and Ψ with the estimator $\widehat{\Psi}_{pq}$ with kernel (4.1). This leads to the equation

$$\sum_{j=1}^{q} \hat{\zeta}_{nj} \hat{u}_j = \sum_{i=1}^{p} \hat{\xi}_{ni} \widehat{\Psi}_{pq}(\hat{v}_i) + \hat{z}_n,$$

where, similarly as in Section 3, $\hat{z}_n$ contains the ε_n, the effect of replacing the infinite sums with finite ones, and the effect of the estimation of the eigenfunctions. This method is based on the residuals defined by

$$\hat{z}_n = \hat{z}_n(p,q) = \sum_{j=1}^{q} \hat{\zeta}_{nj} \hat{u}_j - \sum_{i=1}^{p} \hat{\xi}_{ni} \widehat{\Psi}_{pq}(\hat{v}_i). \tag{4.7}$$

Since $\widehat{\Psi}_{pq}(\hat{v}_i) = \sum_{j=1}^{q} \hat{\lambda}_i^{-1} \hat{\sigma}_{ij} \hat{u}_j(t)$, we see that

$$\hat{z}_n = \sum_{j=1}^{q} \left(\hat{\zeta}_{nj} - \sum_{i=1}^{p} \hat{\xi}_{ni} \hat{\lambda}_i^{-1} \hat{\sigma}_{ij} \right) \hat{u}_j(t).$$

Next define

$$\hat{Z}_{nj} := \langle \hat{u}_j, \hat{z}_n \rangle = \hat{\zeta}_{nj} - \sum_{i=1}^{p} \hat{\xi}_{ni} \hat{\lambda}_i^{-1} \hat{\sigma}_{ij}$$

and denote by $\widehat{\mathbf{C}}_h$ the $q \times q$ autocovariance matrix with entries

$$\hat{c}_h(k,\ell) = \frac{1}{N} \sum_{n=1}^{N-h} (\hat{Z}_{nk} - \hat{\mu}_Z(k))(\hat{Z}_{n+h,\ell} - \hat{\mu}_Z(\ell)),$$

where $\hat{\mu}_Z(k) = N^{-1} \sum_{n=1}^{N} \hat{Z}_{nk}$. Finally, denote by $\hat{r}_{f,h}(i,j)$ and $\hat{r}_{b,h}(i,j)$ the (i,j) entries, respectively, of $\widehat{\mathbf{C}}_0^{-1}\widehat{\mathbf{C}}_h$ and $\widehat{\mathbf{C}}_h\widehat{\mathbf{C}}_0^{-1}$.

The null hypothesis is rejected if the statistic

$$\hat{Q}_N = N \sum_{h=1}^{H} \sum_{i,j=1}^{q} \hat{r}_{f,h}(i,j) \hat{r}_{b,h}(i,j) \tag{4.8}$$

exceeds an upper quantile of the chi-squared distribution with q^2H degrees of freedom; see Theorem 3.

Repeating the arguments in the discussion of Method I, we get the following equivalent expressions for $\hat{Q}_N$:

$$\hat{Q}_N = N \sum_{h=1}^{H} \operatorname{tr}[\widehat{\mathbf{C}}_0^{-1} \widehat{\mathbf{C}}_h^T \widehat{\mathbf{C}}_0^{-1} \widehat{\mathbf{C}}_h]$$

and

$$\hat{Q}_N = N \sum_{h=1}^{H} [\operatorname{vec}(\widehat{\mathbf{C}}_h)]^T [\widehat{\mathbf{C}}_0 \otimes \widehat{\mathbf{C}}_0]^{-1} [\operatorname{vec}(\widehat{\mathbf{C}}_h)].$$

Both methods require the selection of p and q (Method I, only of p). We recommend the popular method based on the cumulative percentage of total variability (CPV) calculated as

$$\operatorname{CPV}(p) = \frac{\sum_{k=1}^{p} \hat{\lambda}_k}{\sum_{k=1}^{\infty} \hat{\lambda}_k},$$

with a corresponding formula for the q. The numbers of eigenfunctions, p and q, are chosen as the smallest numbers, p and q, such that $\operatorname{CPV}(p) \ge 0.85$ and $\operatorname{CPV}(q) \ge 0.85$. Other approaches are available as well, including the scree graph, the pseudo-AIC criterion, BIC, cross-validation, etc. All these methods are implemented in the Matlab PACE package developed at the University of California at Davis.

As p and q increase, the normalized statistics $Q_N^{\wedge}$ and $\hat{Q}_N$ converge to the standard normal distribution. The normal approximation works very well even for small p or q (in the range 3–5 if $N \ge 100$) because the number of the degrees of freedom increases like p^2 or q^2. For Method I, which turns out to be conservative in small samples, the normal approximation brings the size closer to the nominal size. It also improves the power of Method I by up to 10%.

5. ASYMPTOTIC THEORY

The exact asymptotic χ^2-distributions are obtained only under Assumption 1 which, in particular, requires that the X_n be iid. Under Assumptions (A1)–(A5), these χ^2-distributions provide only approximations to the true limit distributions. The approximations are, however, very good, as the simulations in Section 6 show; size and power for dependent X_n are the same as for iid X_n, within the standard error. Thus, to understand the asymptotic properties of the tests, we first consider their behavior under Assumption 1.

Method I is based on the following theorem.

Theorem 2. Suppose Assumptions 1 and 2 and condition (2.1) hold. Then the statistics $Q_N^{\wedge}$ converges to the χ^2-distribution with p^2H degrees of freedom.

Method II is based on Theorem 3. It is analogous to theorem 1 of Gabrys and Kokoszka (2007), but the observations are replaced by residuals (4.7), so a more delicate proof is required.

Theorem 3. Suppose Assumption 1 and condition (2.1) hold. Then statistic (4.8) converges in distribution to a chi-squared random variable with q^2H degrees of freedom.

We now turn to the case of dependent regressors X_n. We focus on Method I. Similar results can be developed to justify the use of Method II, except that the u_j will also be involved. The case of dependent regressors involves the $p \times p$ matrices $\mathbf{D}_h$ with entries

$$D_h(i,j) = \sum_{\ell=p+1}^{\infty} \sum_{k=p+1}^{\infty} \int v_\ell(s) e_h(s,t) v_k(t)\, ds\, dt, \qquad 1 \le i,j \le p,$$

where

$$e_h(s,t) = E[X_0(s) X_h(t)].$$

Theorem 4. Suppose Assumptions (A1)–(A5), 2, and condition (2.1) hold. Then, for any $h > 0$,

$$N^{-1/2}\mathbf{V}_h = N^{-1/2}[\hat{c}_i \hat{c}_j V_h^*(i,j),\ 1 \le i,j \le p] + \mathbf{R}_{N,p}(h) + o_P(1).$$

The matrices $\mathbf{V}_h^* = [V_h^*(i,j),\ 1 \le i,j \le p]$, $1 \le h \le H$, are jointly asymptotically normal. More precisely,

$$N^{-1/2}\{\mathrm{vec}(\mathbf{V}_h^* - N\mathbf{D}_h),\ 1 \le h \le H\} \xrightarrow{d} \{\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_H\},$$

where the p^2-dimensional vectors $\mathbf{Z}_h$ are iid normal, and coincide with the limits of $N^{-1/2}\,\mathrm{vec}(\mathbf{V}_h)$, if the X_n are independent.

For any $r > 0$, the terms $\mathbf{R}_{N,p}(h)$ satisfy,

$$\lim_{p\to\infty} \limsup_{N\to\infty} P\big\{\|\mathbf{R}_{N,p}(h)\| > r\big\} = 0. \tag{5.1}$$

Theorem 4 justifies using Method I for weakly dependent X_n, provided p is so large that the first p FPC v_k explain a large percentage of variance of the X_n. To understand why, first notice that $|D_h(i,j)| \le (\lambda_\ell \lambda_k)^{1/2}$, and since $k, \ell > p$, the eigenvalues λ_ℓ, λ_k are negligible, as for functional datasets encountered in practice the graph of the λ_k approaches zero very rapidly. The exact form of $\mathbf{R}_{N,p}(h)$ can be reconstructed from matrices $\widehat{\mathbf{K}}_p, \widehat{\mathbf{F}}_p, \widehat{\mathbf{G}}_p$ appearing in the *extended version*. If $E[X_0(u)X_h(v)] = 0$, all these matrices (and the matrices $\mathbf{D}_h$) vanish. If the X_n are dependent, these matrices do not vanish, but are negligibly small because they all involve coefficients ψ_{jk} with at least one index greater than p multiplied by factors of order $O_P(N^{-1/2})$. In (5.1), the limit of p increasing to infinity should not be interpreted literally, but again merely indicates that p is so large that the first p FPC v_k explain a large percentage of variance of the X_n.

Our last theorem states conditions under which the test is consistent. The interpretation of the limit as $p \to \infty$ is the same as above. Theorem 5 states that for such p and sufficiently large N the test will reject with large probability if ε_n and ε_{n+h} are correlated in the subspace spanned by $\{v_i,\ 1 \le i \le p\}$.

Theorem 5. Suppose Assumptions (A2), (A4), (A5), (B1)–(B4), 2, and condition (2.1) hold. Then, for all $R > 0$,

$$\lim_{p\to\infty} \liminf_{N\to\infty} P\{Q_N^{\wedge} > R\} = 1,$$

provided $E[\langle \varepsilon_0, v_i\rangle \langle \varepsilon_h, v_j\rangle] \ne 0$, for some $1 \le h \le H$ and $1 \le i,j \le p$.

6. A SIMULATION STUDY

In this section we report the results of a simulation study performed to asses the empirical size and power of the proposed tests (Methods I and II) for small to moderate sample sizes. The sample size N ranges from 50 to 500. Both independent and dependent covariate functions are considered. The simulation runs have 1000 replications each. We used the `R` package `fda`.

For the noise component independent trajectories of the Brownian bridge (BB) and the Brownian motion (BM) are generated by transforming cumulative sums of independent normal random variables computed on a grid of 1000 equispaced points in $[0, 1]$. In order to evaluate the effect of non-Gaussian errors on the finite sample performance, for the noise component we also simulated t_5 and uniform BB and BM (BB_{t_5}, BB_U, BM_{t_5}, and BM_U) by generating t_5 and uniform, instead of normal increments. We also generate errors using Karnhunen–Loéve expansions

$$\varepsilon_n(t) = \sum_{j=1}^{5} \vartheta_{nj} j^{-1/2} \sin(j\pi t),$$

with the iid ϑ_{nj} distributed according to the normal, t_5 and uniform distributions.

We report simulation results obtained using B-spline bases with 20 basis functions, which are suitable for the processes we consider. We also performed the simulations using the Fourier basis and found that they are not significantly different.

To determine the number of principal components (p for X_n and q for Y_n), the cumulative percentage of total variability (CPV) is used as described in Section 4.

Three different kernel functions in (1.1) are considered: the Gaussian kernel

$$\psi(t,s) = \exp\left\{\frac{t^2+s^2}{2}\right\},$$

the Wiener kernel

$$\psi(t,s) = \min(t,s),$$

and the Parabolic kernel

$$\psi(t,s) = -4[(t+1/2)^2 + (s+1/2)^2] + 2.$$

Table 1. Empirical size for independent predictors: $X = \text{BB}$, $\varepsilon = \text{BB}$, $\psi =$ Gaussian, Wiener, and Parabolic, $p = 3$

Sample size	Method I						Method II					
	Gaussian		Wiener		Parabolic		Gaussian		Wiener		Parabolic	
	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
						$H = 1$						
50	6.7	2.5	5.8	3.2	7.4	3.7	7.9	3.7	7.8	3.3	8.2	3.6
100	7.4	3.7	9.5	4.4	8.9	3.8	10.6	5.2	9.9	4.2	9.8	4.7
200	9.8	4.6	8.9	4.2	9.0	4.1	8.9	4.4	10.0	4.0	9.6	4.0
300	9.3	4.8	10.0	5.1	8.1	3.5	8.7	4.4	8.8	4.7	10.3	5.5
500	8.8	5.2	9.8	5.3	9.6	4.9	8.8	4.2	8.9	4.3	8.7	4.0
						$H = 3$						
50	4.3	2.5	5.6	2.1	6.0	3.4	10.7	5.3	8.9	4.7	9.0	4.2
100	7.6	3.7	6.9	3.6	6.4	3.3	9.9	4.5	10.2	4.0	10.1	4.9
200	8.7	4.6	6.4	3.2	8.0	3.3	9.6	4.8	10.1	5.1	9.6	5.0
300	7.6	3.5	9.5	4.2	9.5	4.8	11.0	5.1	8.9	4.0	8.1	4.6
500	9.8	4.6	9.1	3.9	9.2	4.9	11.1	6.8	9.1	4.4	10.0	5.1
						$H = 5$						
50	2.6	0.9	3.5	1.1	4.1	1.4	10.4	5.7	11.2	5.7	10.0	5.1
100	6.5	3.7	5.9	3.0	4.8	1.9	11.3	5.3	10.5	5.2	8.9	4.6
200	8.5	4.4	7.5	3.7	7.4	3.3	11.3	5.7	9.7	4.5	9.7	4.4
300	7.6	4.0	9.9	4.7	7.6	2.8	9.4	4.9	9.8	5.1	10.6	5.5
500	10.1	4.6	9.8	4.4	7.9	3.6	12.1	6.8	9.7	4.7	10.4	5.8

The covariates in (1.1) for both methods are either iid BB or BM, or follow the ARH(1) model of Bosq (2000), which has been extensively used to model weak dependence in functional time series data. To simulate the ARH(1) X_n we used the kernels of the three types above, but multiplied by a constant K, so that their Hilbert–Schmidt norm is 0.5. Thus, the dependent regressors follow the model

$$X_n(t) = K \int_0^1 \psi_X(t, s) X_{n-1}(s)\, ds + \alpha_n(t),$$

where the α_n are iid BB, BM, BB_{t_5}, BB_U, BM_{t_5}, or BM_U.

We present here only a small selection of the results of our numerical experiments. More tables are included in the *extended version*.

Starting with the empirical size, Tables 1 and 2 show that Method I is more conservative and slightly underestimates the nominal levels while Method II tends to overestimate them. The empirical sizes do not depend on whether the BB or the BM is used, nor whether regressors are iid or dependent, nor on the shape of the kernel. These sizes do not deteriorate if errors are not Gaussian either. The empirical size of both methods is thus robust to the form of the kernel, to moderate dependence in the regressors, and to departures from normality in the errors.

For the power simulations, we consider model (1.1) with the Gaussian kernel and $\varepsilon_n \sim \text{ARH}(1)$, that is,

$$\varepsilon_n(t) = K \int_0^1 \psi_\varepsilon(t, s) \varepsilon_{n-1}(s)\, ds + u_n(t),$$

where $\psi_\varepsilon(t, s)$ is Gaussian, Wiener, or Parabolic and K is chosen so that the Hilbert–Schmidt norm of the above ARH(1) operator is 0.5 and the $u_n(t)$ are iid BB, BM, BB_{t_5}, BB_U, BM_{t_5}, or BM_U.

Typical power results are shown in Table 3. Just as for size, power is not affected by the dependence of the regressors. As expected from the results for the empirical size, it is uniformly higher for Method II, but this difference is visible only for $N < 200$ (in our numerical experiments). The power is highest for $H = 1$, especially for smaller samples, because the errors follow the ARH(1) process.

7. APPLICATION TO SPACE PHYSICS AND HIGH–FREQUENCY FINANCIAL DATA

We now illustrate the application of the tests on functional data sets arising in space physics and finance.

Application to Magnetometer Data. Electrical currents flowing in the magnetosphere–ionosphere (M–I) form a complex multiscale system in which a number of individual currents connect and influence each other. Among the various observational means, the global network of ground-based magnetometers stands out with unique strengths of global spacial coverage and real time fine resolution temporal coverage. About a hundred terrestrial geomagnetic observatories form a network, INTERMAGNET, designed to monitor the variations of the M–I current system. Digital magnetometers record three components of the magnetic field in five second resolution, but the INTERMAGNET's data we use consist of one minute averages, that is, 1440 data points per day per component per observatory. Due to the daily rotation of the Earth, we split magnetometer records into days, and treat each daily curve as a single functional observation. We consider the Horizontal (H) component of the magnetic field, lying in the Earth's tangent plane and pointing toward the magnetic North. It most directly reflects the variation of the M–I currents we wish to study.

Table 2. Empirical size for dependent predictors: $X \sim$ ARH(1) with the BB innovations and $\psi_X =$ Gaussian, Wiener, and Parabolic, $\psi =$ Gaussian, $\varepsilon =$ BB, $p = 3$

	Method I						Method II					
	Gaussian		Wiener		Parabolic		Gaussian		Wiener		Parabolic	
Sample size	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
						$H = 1$						
50	8.4	3.9	5.9	2.1	7.3	2.9	9.2	4.6	7.2	2.7	8.6	3.8
100	8.9	4.4	8.8	3.7	8.4	3.7	10.4	4.6	10.2	4.9	9.9	4.8
200	10.2	4.7	9.7	4.6	10.1	4.7	9.5	4.8	8.9	4.0	9.8	5.2
300	9.2	4.9	8.9	4.4	8.6	4.6	10.1	4.1	8.5	3.4	12.0	5.3
500	10.5	5.2	9.3	4.5	9.0	4.7	9.0	4.2	9.5	4.8	11.5	5.6
						$H = 3$						
50	4.4	2.2	5.3	2.9	5.5	2.8	8.1	4.1	10.7	4.5	10.1	4.0
100	6.6	3.1	6.0	2.7	7.0	2.9	10.7	5.4	9.1	4.9	9.9	4.5
200	7.8	3.1	8.5	4.1	8.9	3.9	11.9	6.2	8.5	4.0	7.7	2.9
300	8.2	4.8	8.6	3.9	9.4	4.8	11.9	5.2	8.8	4.4	9.3	5.2
500	11.4	5.3	10.3	5.7	9.1	4.3	10.6	5.4	9.9	5.1	9.9	4.9
						$H = 5$						
50	4.2	1.8	3.2	1.5	4.0	1.9	9.9	5.2	11.1	6.6	11.9	6.7
100	7.2	3.2	4.9	2.4	5.2	2.1	10.5	5.5	10.2	5.5	11.2	6.0
200	7.6	2.8	8.1	3.7	8.8	4.4	11.4	4.6	10.3	4.6	11.6	7.3
300	8.3	4.2	8.3	3.4	7.3	3.9	10.7	5.5	9.3	5.2	9.7	4.7
500	10.7	5.8	10.4	4.9	7.9	4.2	9.0	4.1	9.2	4.0	10.4	5.3

The problem that motivated the examples in this section is that of the association between the auroral (high latitude) electrical currents and the currents flowing at mid and low latitudes. This problem was studied in Maslova et al. (2010), who provide more extensive references to the relevant space physics literature, and to a smaller extent in Kokoszka et al. (2008) and Horváth, Kokoszka, and Reimherr (2009). The problem has been cast into the setting of the functional linear model (1.1) in which the X_n are centered high-latitude records and Y_n are centered mid or low-latitude magnetometer records. We consider two settings: (1) consecutive days and (2) nonconsecutive days on which disturbances known as substorms occur. For consecutive days, we expect the rejection of the null hypothesis as there is a visible dependence of the responses from one day to the next; see the bottom panel of Figure 2. The low-latitude curves, like those measured at Honolulu, exhibit changes on scales of several days. The high-latitude curves exhibit much shorter dependence essentially confined to one day. This is because the auroral electrojects change on a scale of about 4 hours. In setting (2), the answer is less clear: the substorm days are chronologically arranged, but substorms may be separated by several days, and after each substorm the auroral current system resets itself to a quiet state.

Table 3. Method I: Empirical power for dependent predictor functions: $X \sim$ ARH(1) and $\varepsilon \sim$ ARH(1) with the BB innovations, $\psi_\varepsilon = \psi_X =$ Gaussian, Wiener, and Parabolic, $\psi =$ Gaussian, $p = 3$

	Gaussian		Wiener		Parabolic	
Sample size	10%	5%	10%	5%	10%	5%
			$H = 1$			
50	79.2	68.6	68.5	54.0	62.3	47.3
100	99.9	99.6	98.6	96.7	97.7	96.0
200	100	100	100	100	100	100
300	100	100	100	100	100	100
500	100	100	100	100	100	100
			$H = 3$			
50	53.8	40.7	45.4	32.8	40.0	29.0
100	98.0	95.7	93.6	89.5	87.5	81.3
200	100	100	100	99.9	100	99.8
300	100	100	100	100	100	100
500	100	100	100	100	100	100
			$H = 5$			
50	41.2	27.9	31.7	20.8	25.4	15.6
100	95.1	90.3	84.4	74.9	78.2	68.1
200	100	100	100	99.8	99.9	99.3
300	100	100	100	100	100	100
500	100	100	100	100	100	100

To apply the tests, we converted the data to functional objects using 20 spline basis functions, and computed the EFPC's $\hat{v}_k$ and $\hat{u}_j$. For low-latitude magnetometer data, 2 or 3 FPC's are needed to explain 87–89%, or 92–94%, of variability while for high-latitude stations to explain 88–91% of variability we need 8–9 FPC's.

Setting (1) (Consecutive days). We applied both methods to pairs (X_n, Y_n) in which the X_n are daily records at College, Alaska, and the Y_n are the corresponding records at six equatorial stations. Ten such pairs are shown in Figure 2. The samples consisted of all days in 2001, and of about 90 days corresponding to the four seasons. For all six stations and for the whole year the p-values were smaller than 10^{-12}. For the four seasons, all p-values, except two, were smaller than 2%. The higher p-values for the samples restricted to 90 days, are likely due to a smaller seasonal effect (the structure of the M–I system in the northern hemisphere changes with season). We conclude that it

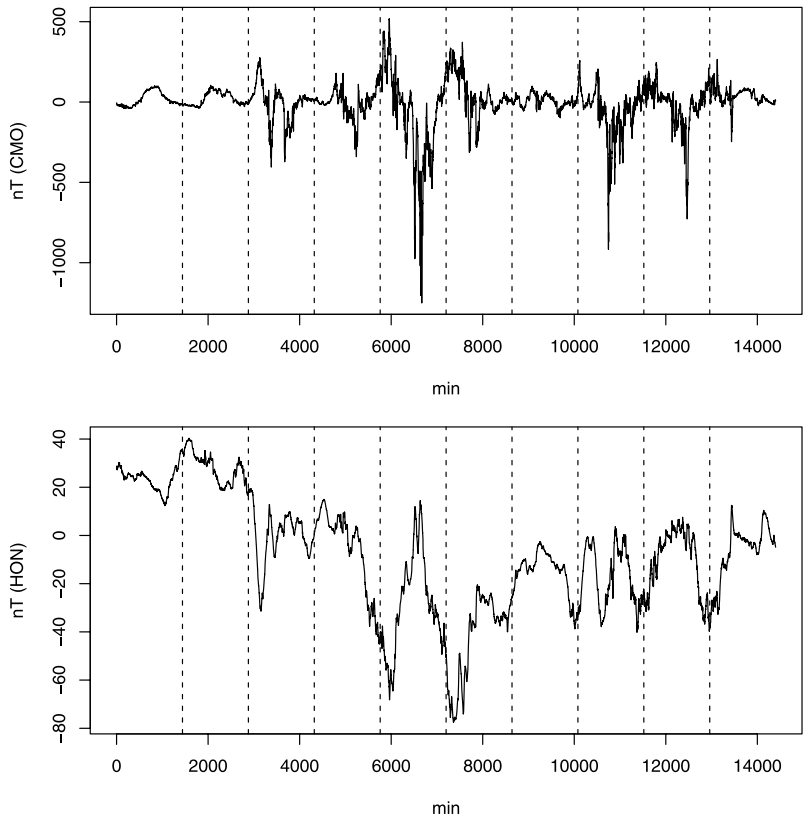

Figure 2. Magnetometer data on 10 consecutive days (separated by vertical dashed lines) recorded at College, Alaska (CMO) and Honolulu, Hawaii (HON).

is not appropriate to use model (1.1) with iid errors to study the interaction of high and low-latitude currents when the data are derived from consecutive days.

Setting (2) (Substorm days). We now focus on two samples studied in Maslova et al. (2010). They are derived from 37 days on which isolated substorms were recorded at College, Alaska (CMO). A substorm is classified as an isolated substorm if it is followed by 2 quiet days. There were only 37 isolated substorms in 2001, data for 10 such days are shown in Figure 3. The first sample consists of 37 pairs (X_n, Y_n), where X_n is the curve of the nth isolated storm recorded at CMO, and Y_n is the curve recorded on the same UT day at Honolulu, Hawai (HON). The second sample is constructed in the same way, except that Y_n is the curve recorded at Boulder, Colorado (BOU). The Boulder observatory is located in geomagnetic midlatitude, that is, roughly halfway between the magnetic north pole and the magnetic equator. Honolulu is located very close to the magnetic equator.

The p-values for both methods and the two samples are listed in Table 4. For Honolulu, both tests indicate the suitability of model (1.1) with iid errors. For Boulder, the picture is less clear. The acceptance by Method I may be due to the small sample size ($N = 37$). The simulations in Section 6 show that for $N = 50$ this method has the power of about 50% at the nominal level of 5%. On the other hand, Method II has the tendency to overreject. The sample with the Boulder records as responses confirms the general behavior of the two methods observed in Section 6, and emphasizes that it is useful to apply both of them to obtain more reliable conclusions. From the space physics perspective, midlatitude records are very difficult to interpret because they combine features of high-latitude events (exceptionally strong auroras have been seen as far south as Virginia) and those of low latitude and field aligned currents.

We also applied the tests to samples in which the regressors are curves on days on which different types of substorms occurred, and the responses are the corresponding curves at low altitude stations. The general conclusion is that for substorm days, the errors in model (1.1) can be assumed iid if the period under consideration is not longer than a few months. For longer periods, seasonal trends cause differences in distribution.

Application to intradaily returns. Perhaps the best known application of linear regression to financial data is the celebrated Capital Asset Pricing Model (CAMP); see, for example, chapter 5 of Campbell, Lo, and MacKinlay (1997). In its simplest form, it is defined by

$$r_n = \alpha + \beta r_{m,n} + \varepsilon_n,$$

where

$$r_n = 100(\ln P_n - \ln P_{n-1}) \approx 100\frac{P_n - P_{n-1}}{P_{n-1}}$$

is the return, in percent, over a unit of time on a specific asset, for example, a stock of a corporation, and $r_{m,n}$ is the analogously defined return on a relevant market index. The unit of time can be can be day, month, or year.

In this section we work with intradaily price data, which are known to have properties quite different than those of daily or

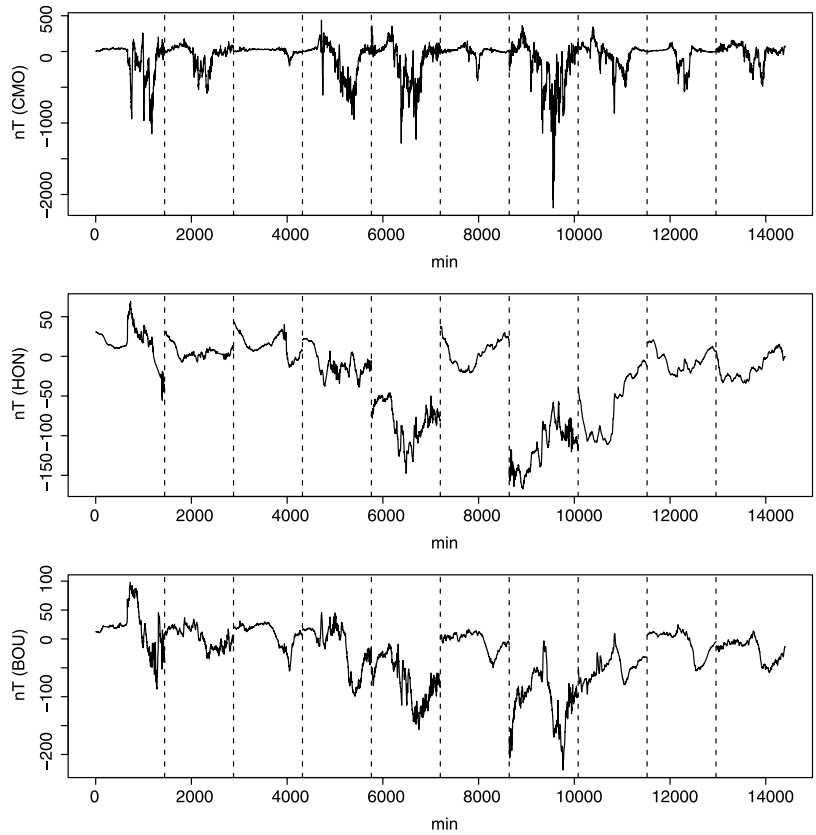

Figure 3. Magnetometer data on 10 chronologically arranged isolated substorm days recorded at College, Alaska (CMO), Honolulu, Hawaii (HON), and Boulder, Colorado (BOU).

monthly closing prices; see, for example, chapter 5 of Tsay (2005); Guillaume et al. (1997) and Andersen and Bollerslev (1997a, 1997b) also offer interesting perspectives. For these data, $P_n(t_j)$ is the price on day n at tick t_j (time of trade); we do not discuss issues related to the bid-ask spread, which are not relevant to what follows. For such data, it is not appropriate to define returns by looking at price movements between the ticks because that would lead to very noisy trajectories for which the methods discussed in this paper, based on the FPC's, are not appropriate; Johnstone and Lu (2009) explain why principal components cannot be meaningfully estimated for such data. Instead, we adopt the following definition.

Definition 1. Suppose $P_n(t_j), n = 1, \ldots, N, j = 1, \ldots, m$, is the price of a financial asses at time t_j on day n. We call the functions

$$r_n(t_j) = 100[\ln P_n(t_j) - \ln P_n(t_1)], \qquad j = 2, \ldots, m,\ n = 1, \ldots, N,$$

the *intradaily cumulative returns*.

Table 4. Isolated substorms data. p-values in percent

	Response	
Method	HON	BOU
I	9.80	26.3
II	6.57	1.15

Figure 4 shows intradaily cumulative returns on 10 consecutive days for the Standard & Poor's 100 index and the Exxon Mobil corporation. These returns have an appearance amenable to smoothing via FPC's.

We propose an extension of the CAPM to such return by postulating that

$$r_n(t) = \alpha(t) + \int \beta(t, s) r_{m,n}(s)\, ds + \varepsilon_n(t), \qquad t \in [0, 1], \quad (7.1)$$

where the interval $[0, 1]$ is the rescaled trading period (in our examples, 9:30 to 16:00 EST). We refer to model (7.1) as the functional CAPM (FCAPM). As far as we know, this model has not been considered in the financial literature, but just as for the classical CAPM, it is designed to evaluate the extent to which intradaily market returns determine the intradaily returns on a specific asset. It is not our goal in this example to systematically estimate the parameters in (7.1) and compare them for various assets and markets, we merely want to use the methods developed in this paper to see if this model can be assumed to hold for some well-known asset. With this goal in mind, we considered FCAPM for S&P 100 and its major component, the Exxon Mobil corporation (currently it contributes 6.78% to this index). The price processes over the period of about 8 years are shown in Figure 5. The functional observations are however not these processes, but the cumulative intradaily returns, examples of which are shown in Figure 4.

After some initial data cleaning and preprocessing steps, we could compute the p-values for any period within the time stretch shown in Figure 5. The p-values for calendar years, the

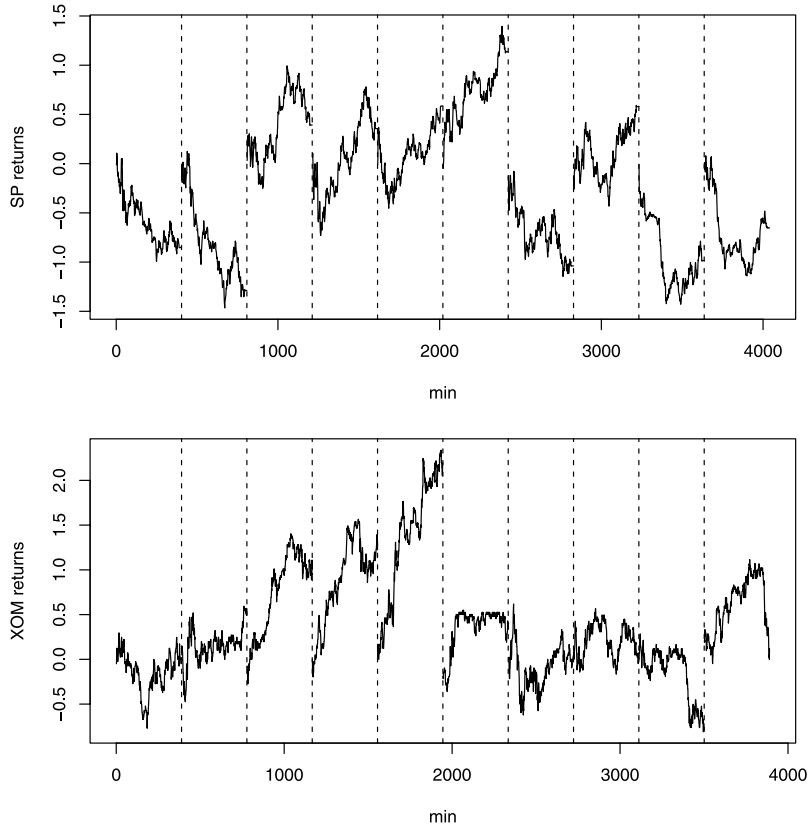

Figure 4. Intradaily cumulative returns on 10 consecutive days for the Standard & Poor's 100 index (S&P) and the Exxon Mobil corporation (XOM).

sample size N is equal to about 250, are reported in Table 5. In this example, both methods lead to the same conclusions, which match the well-known macroeconomic background. The tests do not indicate departures from the FCAMP model, except in 2002, the year between the September 11 attacks and the invasion of Iraq, and in 2006 and 2007, the years preceding the collapse of 2008 in which oil prices were growing at a much faster rate than then the rest of the economy.

In the above examples we tested the correlation of errors in model (1.1). A special case of this model is the historical functional model of Malfait and Ramsay (2003), that is, model (1.1) with $\psi(t,s) = \beta(s,t)I_H(s,t)$, where $\beta(\cdot,\cdot)$ is an arbitrary Hilbert–Schmidt kernel and $I_H(\cdot,\cdot)$ is the indicator function of the set $H = \{(s,t) : 0 \le s \le t \le 1\}$. This model requires that $Y_n(t)$ depends only on the values of $Y_n(s)$ for $s \le t$, that is, it postulates temporal causality within the pairs of curves. Our approach cannot be readily extended to test for error correlation in the historical model of Malfait and Ramsay (2003) because it uses series expansions of a general kernel $\psi(t,s)$, and the restriction that the kernel vanishes in the complement of H does not translate to any obvious restrictions on the coefficients of these expansions. We note, however, that the magnetometer

Figure 5. Share prices of the Standard & Poor's 100 index (S&P) and the Exxon Mobil corporation (XOM). Dashed lines separate years.

Table 5. p-values, in percent, for the FCAPM (7.1) in which the regressors are the intradaily cumulative returns on the Standard & Poor's 100 index, and the responses are such returns on the Exxon Mobil stock

Year	Method I	Method II
2000	46.30	55.65
2001	43.23	56.25
2002	0.72	0.59
2003	22.99	27.19
2004	83.05	68.52
2005	21.45	23.67
2006	2.91	3.04
2007	0.78	0.72

data are obtained at locations with different local times, and for space physics applications the dependence between the shapes of the daily curves is of importance. Temporal causality for financial data is often not assumed as asset values reflect both historical returns and expectations of future market conditions.

[Received December 2009. Revised April 2010.]

REFERENCES

Andersen, T. G., and Bollerslev, T. (1997a), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long Run in High Frequency Data," *Journal of Finance*, 52, 975–1005. [1123]

——— (1997b), "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance*, 2–3, 115–158. [1123]

Aue, A., Hörmann, S., Horváth, L., and Hušková, M. (2010), "Sequential Stability Test for Functional Linear Models," technical report, University of California, Davis. [1115-1117]

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009), "Break Detection in the Covariance Structure of Multivariate Time Series Models," *The Annals of Statistics*, 37, 4046–4087. [1115]

Berkes, I., Hörmann, S., and Schauer, J. (2009), "Asymptotic Results for the Empirical Process of Stationary Sequences," *Stochastic Processes and Their Applications*, 119, 1298–1324. [1115]

Bosq, D. (2000), *Linear Processes in Function Spaces*, New York: Springer. [1114,1116,1120]

Bowerman, B. L., and O'Connell, R. T. (1990), *Linear Statistical Models: An Applied Approach* (2nd ed.), New York: Duxbury. [1113]

Cai, T., and Hall, P. (2006), "Prediction in Functional Linear Regression," *The Annals of Statistics*, 34, 2159–2179. [1113]

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press. [1122]

Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003), "Testing Hypothesis in the Functional Linear Model," *Scandinavian Journal of Statistics*, 30, 241–255. [1113]

Chatfield, C. (1998), "Durbin–Watson Test," in *Encyclopedia of Biostatistics*, eds. P. Armitage and T. Colton, Vol. 2, New York: Wiley, pp. 1252–1253. [1114]

Chiou, J.-M., and Müller, H.-G. (2007), "Diagnostics for Functional Regression via Residual Processes," *Computational Statistics and Data Analysis*, 15, 4849–4863. [1113,1114]

Chiou, J.-M., Müller, H.-G., and Wang, J.-L. (2004), "Functional Response Models," *Statistica Sinica*, 14, 675–693. [1113]

Cochrane, D., and Orcutt, G. H. (1949), "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms," *Journal of the American Statistical Association*, 44, 32–61. [1113]

Cook, D. R. (1977), "Detection of Influential Observations in Linear Regression," *Technometrics*, 19, 15–18. [1113]

Cook, R. D., and Weisberg, S. (1982), *Residuals and Inference in Regression*, New York: Chpaman & Hall. [1113]

Cuevas, A., Febrero, M., and Fraiman, R. (2002), "Linear Functional Regression: The Case of Fixed Design and Functional Response," *The Canadian Journal of Statistics*, 30, 285–300. [1113]

Durbin, J., and Watson, G. S. (1950), "Testing for Serial Correlation in Least Squares Regression. I," *Biometrika*, 37, 409–428. [1114]

——— (1951), "Testing for Serial Correlation in Least Squares Regression. II," *Biometrika*, 38, 159–178. [1114]

——— (1971), "Testing for Serial Correlation in Least Squares Regression. III," *Biometrika*, 58, 1–19. [1114]

Gabrys, R., and Kokoszka, P. (2007), "Portmanteau Test of Independence for Functional Observations," *Journal of the American Statistical Association*, 102, 1338–1348. [1119]

Guillaume, D. M., Dacorogna, M. M., Dave, R. D., Müller, U. A., Olsen, R. B., and Pictet, O. V. (1997), "From the Bird's Eye to the Microscope: A Survey of New Stylized Facts of the Intra-Daily Foreign Exchange Markets," *Finance and Stochastics*, 1, 95–129. [1123]

Hall, P., and Hosseini-Nasab, M. (2006), "On Properties of Functional Principal Components," *Journal of the Royal Statistical Society, Ser. B*, 68, 109–126. [1117]

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57 (2), 357–384. [1113]

Hörmann, S. (2008), "Augmented GARCH Sequences: Dependence Structure and Asymptotics," *Bernoulli*, 14, 543–561. [1115]

Hörmann, S., and Kokoszka, P. (2010), "Weakly Dependent Functional Data," *The Annals of Statistics*, 38, 1845–1884. [1115,1116]

Horn, R. A., and Johnson, C. R. (1991), *Topics in Matrix Analysis*, Cambridge: Cambridge University Press. [1118]

Horváth, L., Kokoszka, P., and Maslova, I. (2009), "Change Point Detection in the Functional Linear Model," technical report, Utah State University. [1115]

Horváth, L., Kokoszka, P., and Reimherr, M. (2009), "Two Sample Inference in Functional Linear Models," *Canadian Journal of Statistics*, 37, 571–591. [1121]

Johnstone, I. M., and Lu, A. Y. (2009), "On Consistency and Sparcity for Principal Components Analysis in High Dimensions," *Journal of the American Statistical Association*, 104, 682–693. [1123]

Kokoszka, P., Maslova, I., Sojka, J., and Zhu, L. (2008), "Testing for Lack of Dependence in the Functional Linear Model," *Canadian Journal of Statistics*, 36, 207–222. [1121]

Li, Y., and Hsing, T. (2007), "On Rates of Convergence in Functional Linear Regression," *Journal of Multivariate Analysis*, 98, 1782–1804. [1113]

Malfait, N., and Ramsay, J. O. (2003), "The Historical Functional Model," *Canadian Journal of Statistics*, 31, 115–128. [1113,1124]

Maslova, I., Kokoszka, P., Sojka, J., and Zhu, L. (2010), "Statistical Significance Testing for the Association of Magnetometer Records at High-, Mid- and Low Latitudes During Substorm Days," *Planetary and Space Science*, 58, 437–445. [1121,1122]

Müller, H.-G., and Stadtmüller, U. (2005), "Generalized Functional Linear Models," *The Annals of Statistics*, 33, 774–805. [1113]

Panaretos, V. M., Kraus, D., and Maddocks, J. H. (2010), "Second-Order Comparison of Gaussian Random Functions and the Geometry of DNA Minicircles," *Journal of the American Statistical Association*, 105, 670–682. [1115]

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis*, New York: Springer. [1113,1116]

Rao, P., and Griliches, Z. (1969), "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors," *Journal of the American Statistical Association*, 64, 253–272. [1113]

Reiss, P. T., and Ogden, R. T. (2007), "Functional Principal Component Regression and Functional Partial Least Squares," *Journal of the American Statistical Association*, 102, 984–996. [1113]

——— (2009), "Smoothing Parameter Selection for a Class of Semiparametric Linear Models," *Journal of the Royal Statistical Society, Ser. B*, 71, 505–523. [1113]

——— (2010), "Functional Generalized Linear Models With Images as Predictors," *Biometrics*, 66, 61–69. [1113]

Sacks, J., and Ylvisaker, D. (1966), "Designs for Regression Problems With Correlated Errors," *The Annals of Mathematical Statistics*, 37, 66–89. [1113]

Schmoyer, R. L. (1994), "Permutations Test for Correlation in Regression Errors," *Journal of the American Statistical Association*, 89, 1507–1516. [1114]

Seber, G. A. F., and Lee, A. J. (2003), *Linear Regression Analysis*, New York: Wiley. [1113,1114]

Thiel, H. (1965), "The Analysis of Disturbances in Regression Analysis," *Journal of the American Statistical Association*, 60, 1067–1079. [1114]

Thiel, H., and Nagar, A. L. (1961), "Testing the Independence of Regression Disturbances," *Journal of the American Statistical Association*, 57, 793–806. [1114]

Tsay, R. S. (2005), *Analysis of Financial Time Series*, New York: Wiley. [1123]

Yao, F., Müller, H.-G., and Wang, J.-L. (2005a), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statististical Association*, 100, 577–590. [1113]

——— (2005b), "Functional Linear Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903. [1113,1116,1117]