Tests for error correlation in the functional linear model

Robertas Gabrys
Utah State University

Lajos Horváth
University of Utah

Piotr Kokoszka
Utah State University

Abstract

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, estimation of the regression kernel.

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 *et al.* (2002), Malfait and Ramsay (2003), Cardot *et al.* (2003), Chiou *et al.* (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, 2009a 2009b), 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

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

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 1.1 shows diagnostic plots of Chiou and Müller

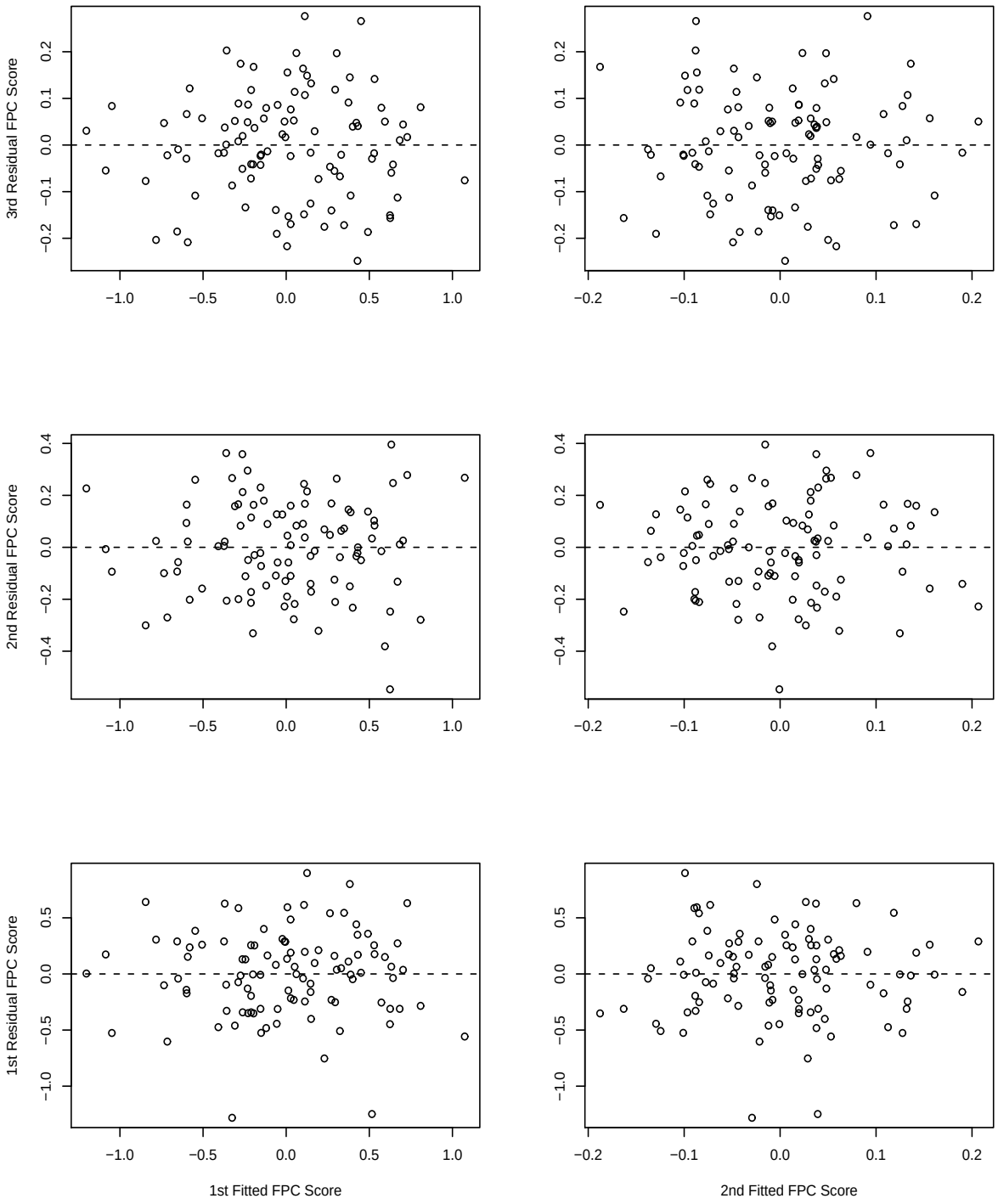
(2007) obtained for synthetic data that follow a functional linear model with highly correlated errors. These plots exhibit almost ideal football shapes. It is equally easy to construct examples in which our methodology fails to detect departures from model (1.1), but the graphs of Chiou and Müller (2007) 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 *et al.* (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 *et al.* (2009). Method II uses two bases: the eigenfunctions of

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



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 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 Sections 8, 9 and 10.

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 2.1 *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 *et al.* (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 2.1; the weak dependence of the $\{X_n\}$ is quantified in Conditions (A2) and (A5). Assumption 2.1 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

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

where

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

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 heteroskedastic 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

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

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

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 $EX_n(t) = 0$)

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

and analogously defined $\hat\Gamma$. The empirical scores are also denoted with the "hat", i.e. 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

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

where

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

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 $v_i,\ i \leq p$, and $u_i,\ i \leq q$.

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), i.e. the independence of the ε_n.

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

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

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 2.1, and in the following, we set

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

THEOREM 2.1 *Suppose Assumptions (A2), (A4) and (A5) hold, and*

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

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

$$\limsup_{N\to\infty} NE\left[||\hat{c}_j\hat{v}_j - v_j||^2\right] < \infty, \qquad \limsup_{N\to\infty} NE\left[|\lambda_j - \hat{\lambda}_j|^2\right] < \infty. \tag{2.2}$$

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, e.g in Ramsay and Silverman (2005) and Yao *et al.* (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, e.g. $Y_{nk} = \langle Y_n, v_k \rangle$ 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 $\psi_{ij} = \iint \psi(t,s) v_i(t) v_j(s) dt ds$. 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

$$\mathbf{X}_n = [\xi_{n1}, \ldots, \xi_{np}]^T \qquad \mathbf{Y}_n = [Y_{n1}, \ldots, Y_{np}]^T, \qquad \boldsymbol{\delta}_n = [\delta_{n1}, \ldots, \delta_{np}]^T,$$

$$\boldsymbol{\psi} = [\psi_{11}, \ldots, \psi_{1p}, \psi_{21}, \ldots, \psi_{2p} \ldots, \psi_{p1}, \ldots, \psi_{pp}]^T,$$

we rewrite (3.2) as

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

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

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

with the "estimator"

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

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

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

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

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

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 $N^{-1/2}\sum_{n=1}^{N} X_n(t)$ 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 $\widetilde{\boldsymbol{\psi}}^{\wedge} - \widetilde{\boldsymbol{\psi}}$ does not follow due to the random signs $\hat{c}_j$. It does however follow from (3.7) that $N^{1/2}(\widetilde{\boldsymbol{\psi}}^{\wedge} - \widetilde{\boldsymbol{\psi}}) = O_P(1)$, and this relation does not require the regularity assumptions of Hall and Hosseini-Nasab (2006). The rate $N^{1/2}$ is optimal, i.e. if $a_N/N^{1/2} \to \infty$, then $a_N(\widetilde{\boldsymbol{\psi}}^{\wedge} - \widetilde{\boldsymbol{\psi}}) \xrightarrow{P} \infty$. This is exactly the result that will be used in the following, and we state it here as Proposition 3.1. We need the following additional assumption.

ASSUMPTION 3.1 *The coefficients ψ_{ij} of the kernel $\psi(\cdot,\cdot)$ satisfy $\sum_{i,j=1}^{\infty}|\psi_{ij}| < \infty$.*

PROPOSITION 3.1 *If Assumptions (A1)–(A5) and 3.1 hold, then $\widetilde{\boldsymbol{\psi}}^{\wedge} - \widetilde{\boldsymbol{\psi}} = O_P(N^{-1/2})$.*

The proof of Proposition 3.1 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 $\widetilde{\boldsymbol{\psi}}^{\wedge}$ 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 $\widetilde{\boldsymbol{\psi}}^{\wedge}$ 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 *et al.* (2005b), we approximate $\psi(\cdot,\cdot)$ by

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

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 $\widetilde{\boldsymbol{\Psi}}_p^{\wedge}$ (3.6) whose (i,j) entry approximates $\hat{c}_i\psi_{ij}\hat{c}_j$, and define also $p \times 1$ vectors

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

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

The fitted vectors are then

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

and the residuals are $\mathbf{R}_n = \hat{\mathbf{Y}}_n - \widetilde{\mathbf{Y}}_n^{\wedge}$. 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 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

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

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

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

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 5.1.

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. Method II 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

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

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 5.2.

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

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

and

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

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

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

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 $CPV(p) \geq 0.85$ and $CPV(q) \geq 0.85$. 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 $N \geq 100$) 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 2.1 which, in particular, requires that the X_n be iid. Under Assumption (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 2.1.

Method I is based on the following theorem which is proven in Section 8.

THEOREM 5.1 *Suppose Assumptions 2.1 and 3.1 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 5.2 which is proven in Section 9. 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 5.2 *Suppose Assumption 2.1 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 $\boldsymbol{D}_h$ with entries

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

where

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

THEOREM 5.3 *Suppose Assumptions (A1)–(A5), Assumption 3.1 and condition* (2.1) *hold. Then, for any* $h > 0$,

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

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

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

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,*

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

Theorem 5.3, proven in Section 10, 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 data sets 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 Lemmas 10.1– 10.3. If $E[X_0(u)X_h(v)] = 0$, all these matrices (and the matrices $\boldsymbol{D}_h$) 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 5.4 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 5.4 *Suppose Assumptions (B1)–(B4), (A2), (A4), (A5), Assumption 3.1 and condition* (2.1) *hold. Then, for all* $R > 0$,

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

provided $E[\langle\varepsilon_0, v_i\rangle\langle\varepsilon_h, v_j\rangle] \neq 0$, *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 (Method I and Method II) for small to moderate sample sizes. Simulations are based on model (1.1). The sample size N takes values ranging from 50 to 500. Both independent and dependent covariate functions are considered. The simulation runs have $1,000$ replications each. The simulations are done in the `R` language, using the `fda` package.

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 $1,000$ 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

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

The first set of runs under H_0 is performed to determine whether for finite sample sizes the procedures achieve nominal 10%, 5%, and 1% levels of significance deduced from the asymptotic distribution. 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.

The empirical rejection rates are collected in Tables 6.1 through 6.8: Method I: Tables 6.1 through 6.4 and Method II: Tables 6.5 through 6.8. The tables show that Method I tends to be more conservative and slightly underestimates the nominal levels while Method II tends to overestimate them, especially for $H = 5$. For samples of size 200 or larger, the procedures achieve significance levels close to the true nominal levels. The tables show that the empirical sizes do not depend on whether the BB or the BM was 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. This shows that the empirical size of both methods is 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 $\varepsilon_n \sim$ ARH(1), i.e.

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

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. Empirical power for all sample sizes considered in the simulation study and for all departures from the model assumptions is summarized in a series of tables: Method I: Tables 6.9 through 6.11, Method II: Tables 6.12 through 6.14. To conserve space results are presented for $\psi =$ Gaussian and $\psi_\varepsilon = \psi_X =$ Gaussian, Wiener and Parabolic. For Method I, $\varepsilon_n =$BB gives slightly higher power than using the BM. For sample sizes $N = 50$ and 100 Method II dominates Method I, but starting with samples of 200 or larger both methods give very high power for both Gaussian and non-Gaussian innovations. Simulations show that the power is not affected on whether regressors are iid or dependent. From the tables, we observe that the power is highest for lag $H = 1$, especially for smaller samples, because the errors follow the ARH(1) process.

TABLE 6.1 Method I: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB$.

Sample size	$p = 3$			$p = 3$			$p = 3$		
	$\psi =$ Gaussian			$\psi =$ Wiener			$\psi =$ Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	6.7	2.5	0.1	5.8	3.2	0.3	7.4	3.7	0.1
100	7.4	3.7	0.7	9.5	4.4	0.8	8.9	3.8	0.6
200	9.8	4.6	0.9	8.9	4.2	0.4	9.0	4.1	0.5
300	9.3	4.8	1.2	10.0	5.1	0.5	8.1	3.5	0.7
500	8.8	5.2	1.1	9.8	5.3	1.1	9.6	4.9	1.3
	$H = 3$								
50	4.3	2.5	0.1	5.6	2.1	0.5	6.0	3.4	0.2
100	7.6	3.7	0.5	6.9	3.6	0.6	6.4	3.3	0.5
200	8.7	4.6	0.6	6.4	3.2	0.7	8.0	3.3	0.8
300	7.6	3.5	0.7	9.5	4.2	1.2	9.5	4.8	0.5
500	9.8	4.6	1.4	9.1	3.9	0.9	9.2	4.9	0.8
	$H = 5$								
50	2.6	0.9	0.1	3.5	1.1	0.1	4.1	1.4	0.1
100	6.5	3.7	0.8	5.9	3.0	0.6	4.8	1.9	0.1
200	8.5	4.4	1.3	7.5	3.7	0.8	7.4	3.3	0.2
300	7.6	4.0	0.6	9.9	4.7	1.0	7.6	2.8	0.3
500	10.1	4.6	1.0	9.8	4.4	1.1	7.9	3.6	0.3

TABLE 6.2 Method I: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB_{t_5}$.

Sample size	$p = 3$			$p = 3$			$p = 3$		
	$\psi =$ Gaussian			$\psi =$ Wiener			$\psi =$ Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	7.4	3.4	0.2	6.4	2.0	0.0	6.4	2.5	0.2
100	8.7	4.2	0.3	5.8	2.8	0.6	9.1	4.3	1.0
200	8.2	3.2	0.7	9.5	4.2	0.8	8.5	4.1	0.9
300	8.8	4.0	0.5	9.3	4.7	0.6	9.2	5.6	1.1
500	8.6	4.0	0.4	11.0	5.6	1.2	8.8	3.8	1.2
	$H = 3$								
50	3.1	1.7	0.2	4.4	1.4	0.3	4.3	1.2	0.3
100	7.0	3.1	0.8	6.8	2.4	0.2	5.8	3.3	0.2
200	7.3	3.4	1.0	11.0	5.6	1.2	8.0	3.5	0.7
300	10.1	5.0	0.5	8.9	3.4	1.0	9.8	4.0	0.7
500	10.8	6.4	1.1	9.2	5.7	1.2	10.2	5.5	1.2
	$H = 5$								
50	3.8	0.7	0.0	4.4	2.5	1.1	3.5	1.5	0.4
100	5.4	2.4	0.3	4.8	1.9	0.2	5.1	2.7	0.8
200	10.2	4.5	1.0	6.9	3.8	0.8	7.1	3.4	0.5
300	10.1	5.1	1.2	9.4	4.2	0.7	9.3	4.7	0.9
500	10.2	5.1	1.3	10.3	5.1	1.3	8.6	3.7	1.0

TABLE 6.3 Method I: Empirical size for independent regressor functions: $X_n = BB_n$, $\varepsilon_n = \sum_{j=1}^{5} \vartheta_{nj} \cdot j^{-1/2} \cdot \sin(j\pi t)$, $n = 1, \ldots, N$, $\vartheta_{nj} \sim N(0, 1)$.

Sample size	$p = 3$			$p = 3$			$p = 3$		
	$\psi =$ Gaussian			$\psi =$ Wiener			$\psi =$ Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	6.5	1.9	0.2	7.2	2.2	0.1	8.6	2.5	0.4
100	9.1	4.6	0.7	9.0	4.6	0.5	9.5	4.3	0.6
200	9.5	4.8	1.3	8.7	4.2	0.9	9.3	4.8	0.4
300	8.1	3.6	1.2	8.5	4.4	0.7	9.6	4.6	0.9
500	9.3	4.4	0.7	9.5	4.1	0.8	10.3	4.5	0.6
	$H = 3$								
50	4.0	1.2	0.0	5.5	2.0	0.1	4.3	1.4	0.2
100	6.9	3.2	1.0	6.9	3.3	0.6	7.5	3.1	0.7
200	10.0	5.5	0.8	8.5	4.4	0.7	7.7	3.9	1.2
300	10.1	4.7	0.8	8.3	3.5	0.6	7.3	2.9	0.4
500	7.9	4.2	1.0	6.9	2.9	0.6	9.1	4.7	0.6
	$H = 5$								
50	2.9	1.4	0.1	3.8	1.5	0.2	3.3	0.9	0.0
100	5.5	2.2	0.3	4.2	2.3	0.4	5.8	2.4	0.3
200	9.3	4.6	0.5	7.2	3.5	0.5	7.7	2.7	0.4
300	7.1	3.3	0.7	7.2	3.9	0.8	8.0	4.1	0.9
500	9.9	5.0	1.0	9.1	4.2	1.0	8.6	4.0	1.0

TABLE 6.4 Method I: Empirical size for dependent predictors: $X \sim ARH(1)$ with the BB innovations, ψ =Gaussian, $\varepsilon = BB$.

Sample size	$p = 3$			$p = 3$			$p = 3$		
	ψ_X = Gaussian			ψ_X = Wiener			ψ_X = Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	8.4	3.9	0.3	5.9	2.1	0.5	7.3	2.9	0.3
100	8.9	4.4	0.7	8.8	3.7	0.3	8.4	3.7	0.7
200	10.2	4.7	0.9	9.7	4.6	0.5	10.1	4.7	0.9
300	9.2	4.9	0.8	8.9	4.4	0.8	8.6	4.6	0.9
500	10.5	5.2	1.4	9.3	4.5	0.6	9.0	4.7	0.7
	$H = 3$								
50	4.4	2.2	0.3	5.3	2.9	0.4	5.5	2.8	0.3
100	6.6	3.1	0.3	6.0	2.7	0.6	7.0	2.9	0.6
200	7.8	3.1	0.5	8.5	4.1	1.1	8.9	3.9	0.3
300	8.2	4.8	0.7	8.6	3.9	1.1	9.4	4.8	1.2
500	11.4	5.3	1.5	10.3	5.7	1.3	9.1	4.3	0.5
	$H = 5$								
50	4.2	1.8	0.1	3.2	1.5	0.2	4.0	1.9	0.2
100	7.2	3.2	0.6	4.9	2.4	0.7	5.2	2.1	0.4
200	7.6	2.8	0.9	8.1	3.7	1.3	8.8	4.4	1.1
300	8.3	4.2	0.6	8.3	3.4	0.9	7.3	3.9	0.0
500	10.7	5.8	0.9	10.4	4.9	1.3	7.9	4.2	0.9

TABLE 6.5 Method II: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB$.

Sample size	$p = 3$, $q = 2$			$p = 3$, $q = 3$			$p = 3$, $q = 2$		
	$\psi =$ Gaussian			$\psi =$ Wiener			$\psi =$ Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	7.9	3.7	0.4	7.8	3.3	0.7	8.2	3.6	0.4
100	10.6	5.2	1.4	9.9	4.2	0.3	9.8	4.7	0.5
200	8.9	4.4	0.9	10.0	4.0	0.5	9.6	4.0	0.7
300	8.7	4.4	0.5	8.8	4.7	0.4	10.3	5.5	0.9
500	8.8	4.2	1.1	8.9	4.3	1.0	8.7	4.0	0.7
	$H = 3$								
50	10.7	5.3	0.9	8.9	4.7	1.0	9.0	4.2	1.0
100	9.9	4.5	1.0	10.2	4.0	0.5	10.1	4.9	0.6
200	9.6	4.8	0.9	10.1	5.1	0.9	9.6	5.0	0.9
300	11.0	5.1	1.1	8.9	4.0	0.8	8.1	4.6	1.1
500	11.1	6.8	1.3	9.1	4.4	0.6	10.0	5.1	1.4
	$H = 5$								
50	10.4	5.7	1.1	11.2	5.7	1.2	10.0	5.1	1.2
100	11.3	5.3	1.1	10.5	5.2	1.1	8.9	4.6	1.0
200	11.3	5.7	1.1	9.7	4.5	0.8	9.7	4.4	0.8
300	9.4	4.9	0.5	9.8	5.1	0.8	10.6	5.5	0.8
500	12.1	6.8	1.2	9.7	4.7	1.3	10.4	5.8	1.1

TABLE 6.6 Method II: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB_{t_5}$.

Sample size	$p = 3$, $q = 2$			$p = 3$, $q = 3$			$p = 3$, $q = 2$		
	$\psi =$ Gaussian			$\psi =$ Wiener			$\psi =$ Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	8.3	3.8	0.6	7.4	3.4	0.2	9.4	4.5	0.8)
100	9.6	3.7	0.8	9.4	4.1	0.6	10.5	5.0	1.5
200	8.1	3.9	1.0	9.2	5.7	0.8	9.7	4.7	0.7
300	10.7	5.5	1.5	8.6	4.2	0.8	12.2	5.1	0.8
500	11.6	5.6	1.3	8.9	4.2	1.1	10.8	3.9	0.5
	$H = 3$								
50	8.6	3.4	0.5	9.2	4.4	0.5	8.8	4.7	1.2
100	9.7	4.9	0.6	11.1	4.8	0.8	10.5	5.4	0.9
200	8.9	5.6	1.3	10.1	4.4	1.3	9.2	5.2	0.9
300	11.	5.6	1.0	10.5	5.8	1.0	8.5	4.5	0.7
500	10.	6.3	0.8	10.6	5.0	0.7	10.0	4.8	0.4
	$H = 5$								
50	10.9	5.7	1.9	12.6	6.4	1.6	10.6	5.7	1.6
100	10.6	6.0	1.4	10.7	5.4	1.2	10.6	4.7	1.6
200	10.6	6.2	1.2	9.5	4.3	0.5	11.5	5.9	1.2
300	10.5	5.5	1.0	9.9	4.3	0.8	9.4	5.3	1.1
500	10.6	5.0	0.9	9.3	4.7	0.5	9.4	4.6	0.7

TABLE 6.7 Method II: Empirical size for independent regressor functions: $X_n = BB_n$, $\varepsilon_n = \sum_{j=1}^{5} \vartheta_{nj} \cdot j^{-1/2} \cdot \sin(j\pi t)$, $n = 1, \ldots, N$, $\vartheta_{nj} \sim N(0,1)$.

Sample size	$p=3$, $q=4$			$p=3$, $q=4$			$p=3$, $q=4$		
	ψ = Gaussian			ψ = Wiener			ψ = Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	7.2	2.7	0.2	6.8	3.2	0.4	6.9	2.7	0.2
100	9.3	4.6	0.5	7.6	3.8	0.5	7.5	3.7	1.0
200	10.1	4.3	0.5	8.5	4.3	0.9	10.0	4.2	0.8
300	9.2	4.1	0.6	9.7	5.5	0.7	8.8	4.6	1.0
500	9.3	4.3	0.7	10.3	4.5	0.7	8.5	4.3	1.2
	$H = 3$								
50	8.3	3.8	0.5	9.6	4.8	0.4	9.1	3.7	0.3
100	10.0	5.2	0.9	8.5	4.1	0.6	10.3	5.3	0.9
200	9.5	4.0	0.9	10.2	4.9	1.2	9.9	4.5	0.5
300	7.9	4.0	0.8	9.5	4.8	1.4	9.1	4.5	0.9
500	9.6	4.9	1.1	9.4	4.6	0.5	10.1	5.0	0.9
	$H = 5$								
50	13.7	7.0	2.3	12.3	6.5	2.0	12.7	6.2	2.2
100	12.7	6.2	1.1	11.6	5.4	0.7	12.5	5.9	0.7
200	10.7	5.1	0.8	10.9	5.0	1.4	11.2	5.2	1.1
300	10.1	4.5	1.0	9.8	4.0	0.7	10.7	5.4	1.7
500	9.5	4.7	0.6	9.6	4.8	1.3	9.6	4.9	1.3

TABLE 6.8 Method II: Empirical size for dependent predictor functions: $X \sim ARH(1)$ with the BB innovations, ψ =Gaussian, $\varepsilon = BB$.

Sample size	$p=3$			$p=3$			$p=3$		
	ψ_X = Gaussian			ψ_X = Wiener			ψ_X = Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H=1$								
50	9.2	4.6	0.3	7.2	2.7	0.6	8.6	3.8	0.7
100	10.4	4.6	1.0	10.2	4.9	0.7	9.9	4.8	0.6
200	9.5	4.8	1.0	8.9	4.0	0.7	9.8	5.2	0.5
300	10.1	4.1	0.7	8.5	3.4	0.9	12.0	5.3	1.1
500	9.0	4.2	0.8	9.5	4.8	1.2	11.5	5.6	0.6
	$H=3$								
50	8.1	4.1	1.3	10.7	4.5	1.0	10.1	4.0	1.0
100	10.7	5.4	1.0	9.1	4.9	1.1	9.9	4.5	0.8
200	11.9	6.2	1.9	8.5	4.0	0.8	7.7	2.9	0.3
300	11.9	5.2	1.3	8.8	4.4	0.9	9.3	5.2	1.1
500	10.6	5.4	1.2	9.9	5.1	0.6	9.9	4.9	1.4
	$H=5$								
50	9.9	5.2	1.7	11.1	6.6	1.4	11.9	6.7	1.8
100	10.5	5.5	1.2	10.2	5.5	1.0	11.2	6.0	2.2
200	11.4	4.6	0.4	10.3	4.6	1.2	11.6	7.3	1.5
300	10.7	5.5	1.9	9.3	5.2	0.8	9.7	4.7	1.3
500	9.0	4.1	0.8	9.2	4.0	1.0	10.4	5.3	1.3

TABLE 6.9 Method I: Empirical power for independent predictors: $X = BB$, $\varepsilon \sim ARH(1)$ with the BB innovations.

Sample size	$p = 3$			$p = 3$			$p = 3$		
	ψ_ε = Gaussian			ψ_ε = Wiener			ψ_ε = Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	84.3	75.3	46.0	53.7	36.7	12.9	82.8	70.9	43.8
100	99.9	99.7	98.0)	96.1	92.2	76.3	99.8	99.7	98.7
200	100	100	100	100	100	99.7	100	100	100
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 3$								
50	60.4	46.2	24.0	35.7	25.3	9.5	63.4	49.7	26.5
100	97.9	96.9	88.9	83.9	75.7	54.8	97.8	96.4	90.3
200	100	100	100	99.9	99.5	97.6	100	100	100
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 5$								
50	43.2	32.4	15.3	24.5	16.1	6.3	44.2	31.8	15.4
100	94.6	90.5	75.6	72.4	61.5	42.2	95.3	90.0	76.5
200	100	100	99.8	99.2	98.0	94.4	100	100	99.9
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100

TABLE 6.10 Method I: Empirical power for independent predictors: $X = BB$, $\varepsilon \sim ARH(1)$ with the BB_{t_5} innovations.

Sample size	$p = 3$			$p = 3$			$p = 3$		
	ψ_ε = Gaussian			ψ_ε = Wiener			ψ_ε = Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	85.1	73.6	46.1	52.4	37.6	13.7	86.6	75.4	47.5
100	99.7	99.7	98.0	95.5	92.0	76.3	99.9	99.8	98.4
200	100	100	100	100	100	99.8	100	100	100
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 3$								
50	60.7	47.6	24.1	34.8	23.2	9.4	61.5	47.9	26.3
100	98.7	96.5	88.8	83.8	75.7	54.5	99.1	97.9	91.6
200	100	100	100	99.7	99.3	97.3	100	100	100
300	100	100	100	100	100	99.9	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 5$								
50	40.8	29.8	13.8	25.2	16.3	7.0	42.4	29.8	12.8
100	95.0	91.1	76.6	75.6	64.9	42.3	95.8	91.6	79.2
200	100	100	100	99.2	98.6	93.6	100	100	100
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100

TABLE 6.11 Method I: Empirical power for dependent predictor functions: $X \sim ARH(1)$ with the BB innovations, $\varepsilon \sim ARH(1)$ with the BB innovations.

Sample size	$p = 3$			$p = 3$			$p = 3$		
	$\psi_\varepsilon = \psi_X =$ Gaussian			$\psi_\varepsilon = \psi_X =$ Wiener			$\psi_\varepsilon = \psi_X =$ Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	79.2	68.6	40.1	68.5	54.0	26.0	62.3	47.3	20.8
100	99.9	99.6	97.9	98.6	96.7	88.4	97.7	96.0	86.6
200	100	100	100	100	100	100	100	100	100
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 3$								
50	53.8	40.7	19.8	45.4	32.8	14.5	40.0	29.0	13.1
100	98.0	95.7	87.2	93.6	89.5	73.9	87.5	81.3	64.2
200	100	100	100	100	99.9	99.6	100	99.8	99.6
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 5$								
50	41.2	27.9	12.3	31.7	20.8	7.8	25.4	15.6	6.1
100	95.1	90.3	76.4	84.4	74.9	56.1	78.2	68.1	49.0
200	100	100	99.9	100	99.8	99.0	99.9	99.3	97.5
300	100	100	100	100	100	100	100	100	99.9
500	100	100	100	100	100	100	100	100	100

TABLE 6.12 Method II: Empirical power for independent predictors: $X = BB$, $\varepsilon \sim ARH(1)$ with the BB innovations.

Sample size	$p = 3$, $q = 2$			$p = 3$, $q = 3$			$p = 3$, $q = 2$		
	ψ_ε = Gaussian			ψ_ε = Wiener			ψ_ε = Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	86.1	78.6	57.1	86.2	76.9	52.5	80.4	68.2	42.1
100	99.2	98.6	95.4	99.8	99.0	96.9	99.2	98.6	95.2
200	100	100	99.9	100	100	100	100	100	100
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 3$								
50	74.4	63.4	43.1	72.2	61.9	40.8	65.1	52.7	32.1
100	97.6	94.4	89.3	98.7	97.0	91.1	96.2	93.7	86.4)
200	100	100	99.9	100	100	100	100	100	99.7
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 5$								
50	66.3	55.9	34.3	64.0	51.8	32.8	58.1	48.2	26.2
100	95.4	92.6	82.2	96.6	93.8	84.7	93.3	89.6	76.5
200	100	100	99.6	100	100	99.8	100	100	99.5
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100

TABLE 6.13 Method II: Empirical power for independent predictors: $X = BB$, $\varepsilon \sim ARH(1)$ with the BB_{t_5} innovations.

Sample size	$p = 3$, $q = 2$			$p = 3$, $q = 3$			$p = 3$, $q = 2$		
	ψ_ε = Gaussian			ψ_ε = Wiener			ψ_ε = Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	83.2	72.3	50.9	82.5	73.9	47.8	78.2	65.7	40.8
100	99.4	97.3	93.1	99.4	99.1	96.7	99.4	98.5	92.2
200	100	100	100	100	100	100	100	100	99.8
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 3$								
50	70.8	59.8	36.9	68.2	56.9	35.4	64.7	52.0	32.8
100	95.2	92.6	83.9	97.8	95.2	88.6	95.3	91.8	81.9
200	99.9	99.9	99.6	100	100	100	99.9	99.9	99.7
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 5$								
50	62.4	51.2	31.7	63.4	52.4	32.1	54.3	44.4	24.0
100	93.8	88.0	75.1	94.6	91.0	79.5	93.1	87.7	74.0
200	100	99.0	99.4	100	100	99.8	100	100	99.4
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100

TABLE 6.14 Method II: Empirical power for dependent predictors: $X \sim ARH(1)$ with the BB innovations; $\varepsilon = BB \sim ARH(1)$ with the BB innovations.

Sample size	$p = 3$			$p = 3$			$p = 3$		
	$\psi_\varepsilon = \psi_X =$ Gaussian			$\psi_\varepsilon = \psi_X =$ Wiener			$\psi_\varepsilon = \psi_X =$ Parabolic		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
	$H = 1$								
50	86.0	77.5	57.4	85.4	76.1	52.1	79.8	70.0	44.0
100	99.7	98.9	95.5	99.3	99.0	96.9	99.4	98.9	95.0
200	100	100	100	100	100	100	100	100	99.9
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 3$								
50	73.8	61.1	40.4	71.6	60.1	38.9	63.4	50.9	28.7
100	96.8	95.3	90.1	98.8	96.5	90.4	96.9	93.3	82.0
200	99.9	99.9	99.8	100	100	99.9	100	100	99.8
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100
	$H = 5$								
50	65.8	56.2	36.3	64.6	53.1	31.7	59.5	47.4	26.3
100	95.0	91.7	83.3	97.1	93.8	84.2	92.9	87.9	75.3
200	99.9	99.8	99.5	100	100	100	99.8	99.7	99.1
300	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100

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, i.e. 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.

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.* (2009), who provide more extensive references to the relevant space physics literature, and to a smaller extent in Kokoszka *et al.* (2008) and Horváth *et al.* (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, 2) non-consecutive 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 7.1. 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.

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 $92-94$, percent of variability while for high

TABLE 7.1 Isolated substorms data. P-values in percent.

	Response	
Method	HON	BOU
I	9.80	26.3
II	6.57	1.15

latitude stations to explain $88-91$ percent 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 7.1. 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 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.* (2009). 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 7.2. 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, i.e. roughly half way 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 7.1. 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

FIGURE 7.1 Magnetometer data on 10 consecutive days (separated by vertical dashed lines) recorded at College, Alaska (CMO) and Honolulu, Hawaii, (HON).

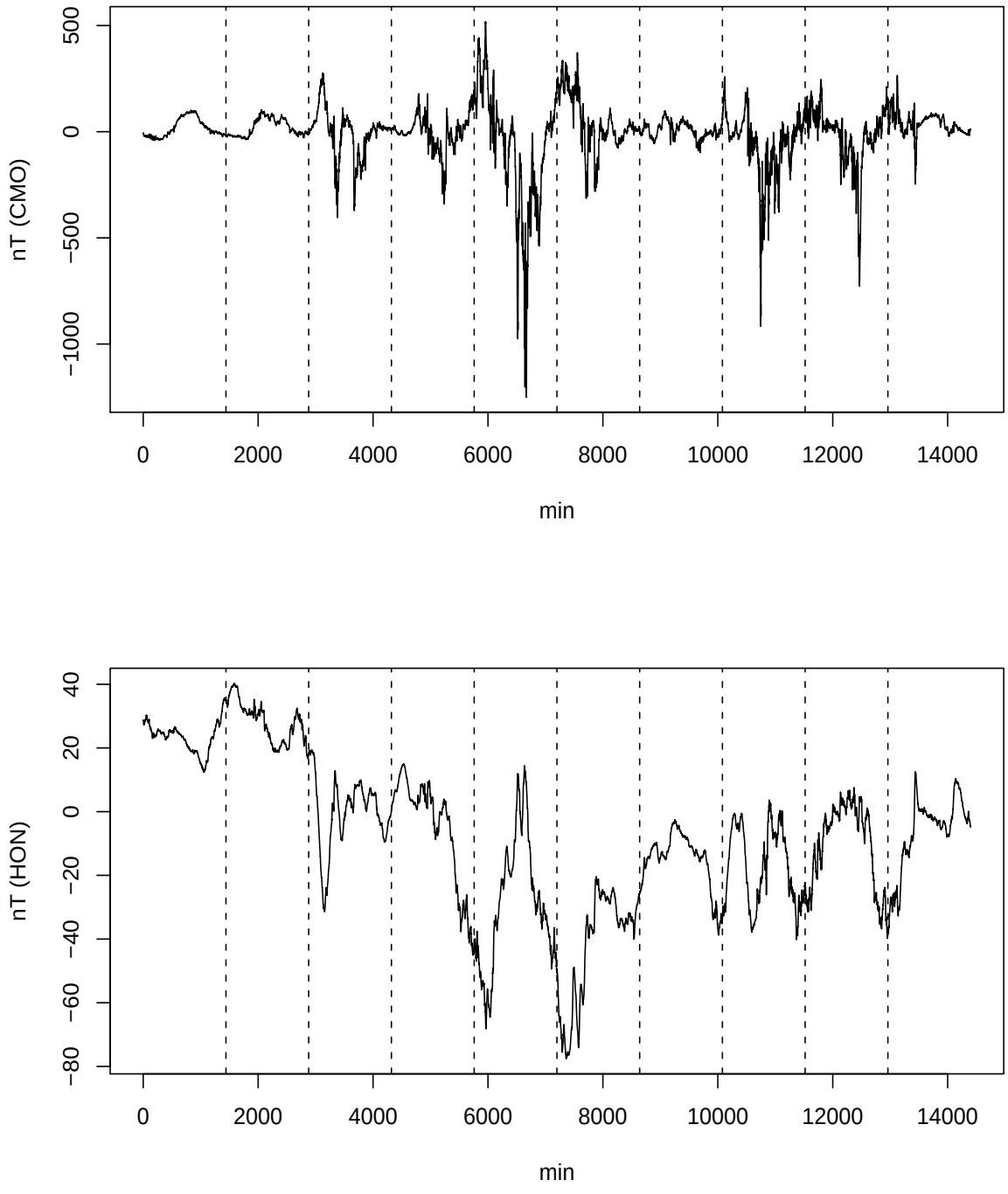
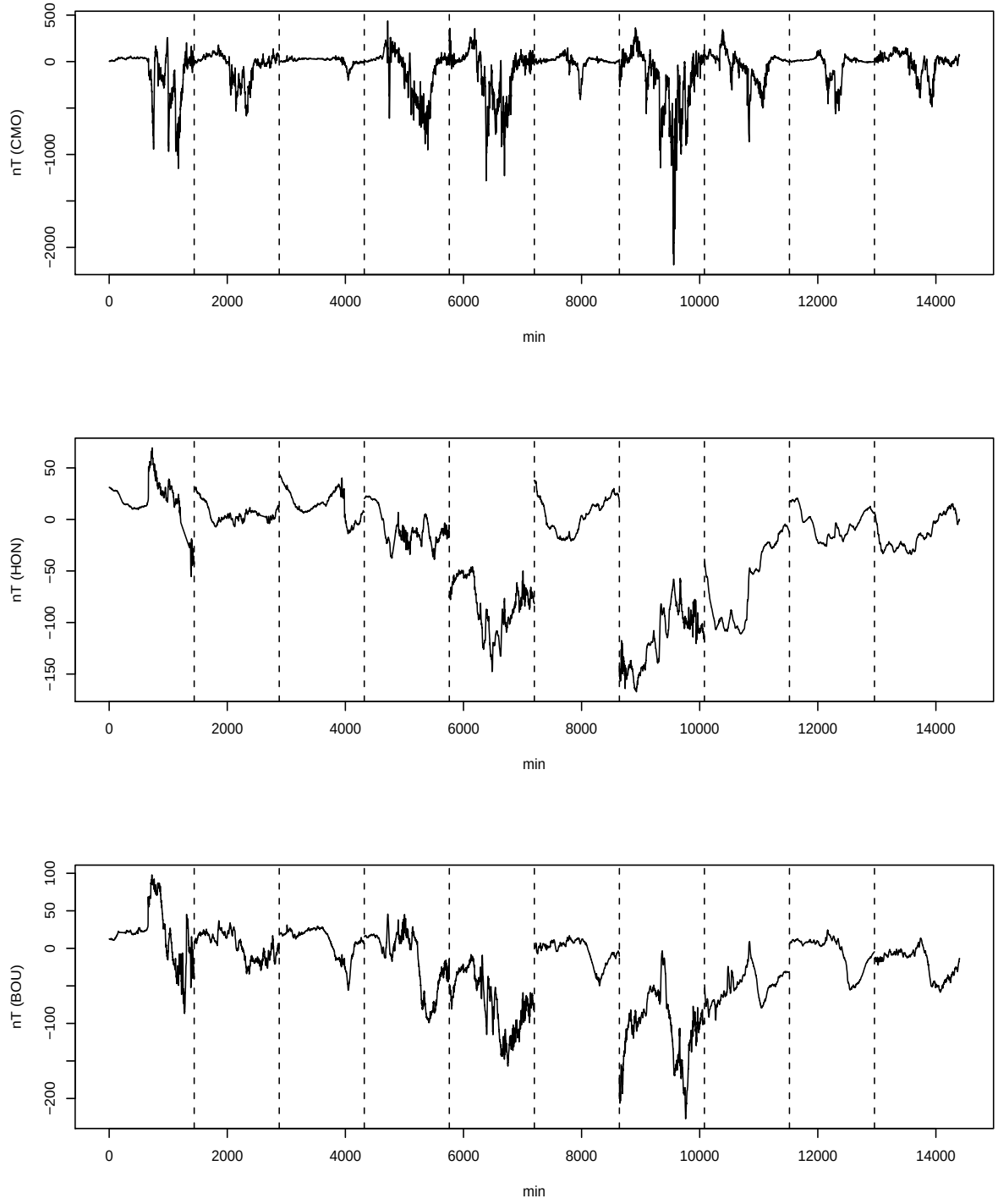


FIGURE 7.2 Magnetometer data on 10 chronologically arranged isolated substorm days recorded at College, Alaska (CMO), Honolulu, Hawaii, (HON) and Boulder, Colorado (BOU).



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 e.g. Chapter 5 of Campbell *et al.* (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, e.g. 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 intra–daily price data, which are known to have properties quite different than those of daily or monthly closing prices, see e.g. 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 7.1 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

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

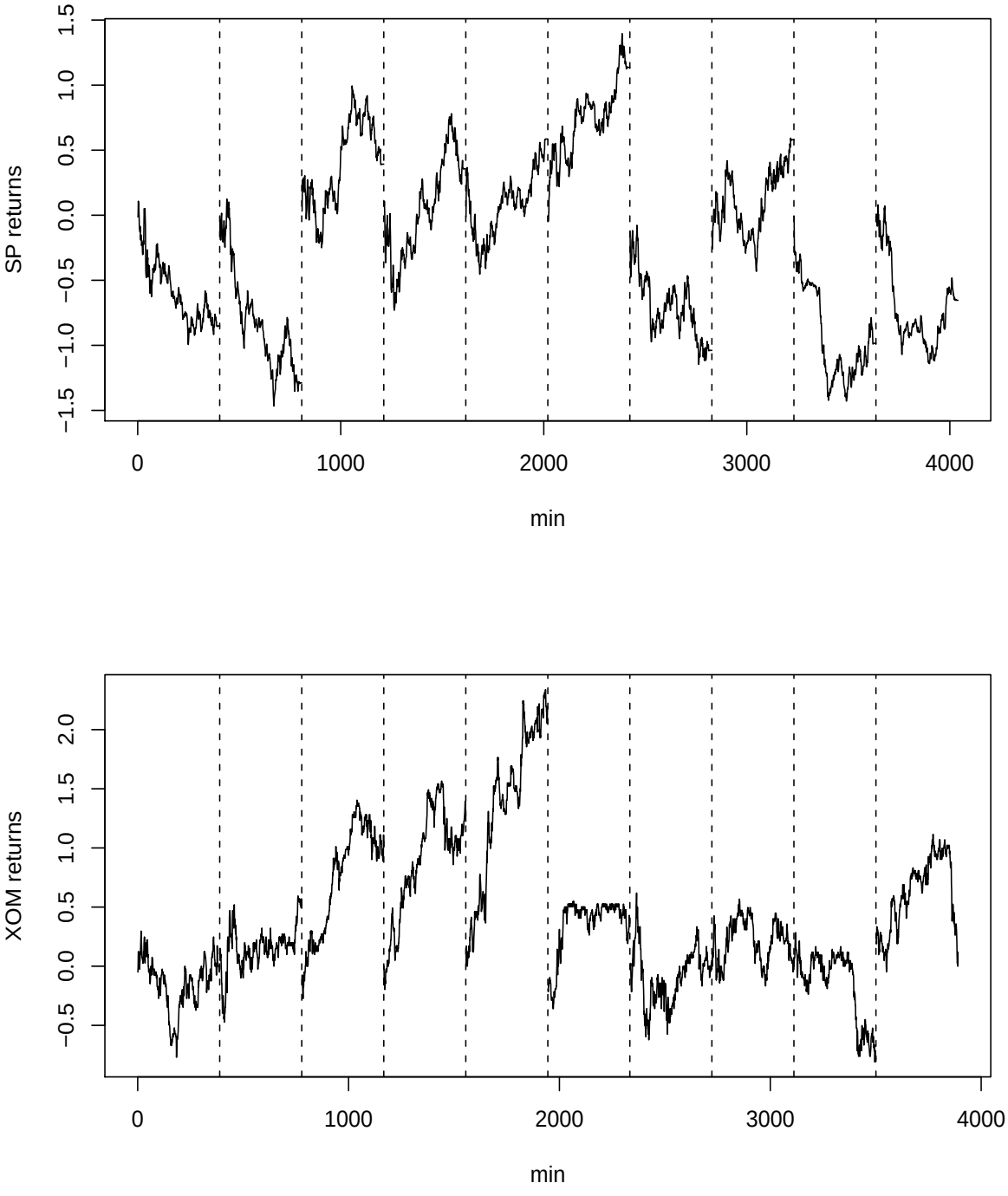
the *intra-daily cumulative returns*.

Figure 7.3 shows intra-daily 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

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

FIGURE 7.3 Intra-daily cumulative returns on 10 consecutive days for the Standard & Poor's 100 index (SP) and the Exxon–Mobil corporation (XOM).



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 intra–daily 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 7.4. The functional observations are however not these processes, but the cumulative intra–daily returns, examples of which are shown in Figure 7.3.

After some initial data cleaning and preprocessing steps, we could compute the p-values for any period within the time stretch shown in Figure 7.4. The p-values for calendar years, the sample size N is equal to about 250, are reported in Table 7.2. 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 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), i.e. 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 $H = \{(s, t) : \ 0 \leq s \leq t \leq 1\}$. This model requires that $Y_n(t)$ depends only on the values of $Y_n(s)$ for $s \leq t$, i.e. 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 data are obtained at locations with different local times, and for space physics aplications 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.

FIGURE 7.4 Share prices of the Standard & Poor's 100 index (SP) and the Exxon–Mobil corporation (XOM). Dashed lines separate years.

TABLE 7.2 P–values, in percent, for the FCAPM (7.1) in which the regressors are the intra–daily 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

8 Proof of Theorem 5.1

Relation (3.2) can be rewritten as

$$\mathbf{Y}_n = \mathbf{\Psi}_p \mathbf{X}_n + \boldsymbol{\delta}_n, \tag{8.1}$$

where

$$\mathbf{\Psi}_p = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1p} \\ \psi_{21} & \psi_{22} & \cdots & \psi_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \psi_{p1} & \psi_{p2} & \cdots & \psi_{pp} \end{bmatrix}.$$

The vectors $\mathbf{Y}_n, \mathbf{X}_n, \boldsymbol{\delta}_n$ are defined in Section 3 as the projections on the FPC's $v_1, v_2, \ldots v_p$. Proposition 8.1 establishes an analog of (8.1) if these FPC's are replaced by the EFPC's $\hat{v}_1, \hat{v}_2, \ldots \hat{v}_p$. These replacement introduces additional terms generically denoted with the letter γ. First we prove Lemma 8.1 which leads to a decomposition analogous to (3.2).

LEMMA 8.1 *If relation* (1.1) *holds with a Hilbert–Schmidt kernel* $\psi(\cdot,\cdot)$*, then*

$$Y_n(t) = \int \left(\sum_{i,j=1}^{p} \hat{c}_i \psi_{ij} \hat{c}_j \hat{v}_i(t) \hat{v}_j(s) \right) X_n(s) ds + \Delta_n(t),$$

where

$$\Delta_n(t) = \varepsilon_n(t) + \eta_n(t) + \gamma_n(t).$$

The terms $\eta_n(t)$ *and* $\gamma_n(t)$ *are defined as follows:*

$$\eta_n(t) = \eta_{n1}(t) + \eta_{n2}(t);$$

$$\eta_{n1}(t) = \int \left(\sum_{i=p+1}^{\infty} \sum_{j=1}^{\infty} \psi_{ij} v_i(t) v_j(s) \right) X_n(s) ds,$$

$$\eta_{n2}(t) = \int \left(\sum_{i=1}^{p} \sum_{j=p+1}^{\infty} \psi_{ij} v_i(t) v_j(s) \right) X_n(s) ds.$$

$$\gamma_n(t) = \gamma_{n1}(t) + \gamma_{n2}(t);$$

$$\gamma_{n1}(t) = \int \sum_{i,j=1}^{p} \hat{c}_i \psi_{ij} [\hat{c}_i v_i(t) - \hat{v}_i(t)] v_j(s) X_n(s) ds,$$

$$\gamma_{n2}(t) = \int \sum_{i,j=1}^{p} \hat{c}_i \psi_{ij} \hat{c}_j \hat{v}_i(t) [\hat{c}_j v_j(s) - \hat{v}_j(s)] X_n(s) ds.$$

PROOF OF LEMMA 8.1. Observe that by (3.1),

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

$$= \int \left(\sum_{i,j=1}^{p} \psi_{ij}v_i(t)v_j(s)\right) X_n(s)ds + \eta_n(t).$$

Thus model (1.1) can be written as

$$Y_n(t) = \int \left(\sum_{i,j=1}^{p} \psi_{ij}v_i(t)v_j(s)\right) X_n(s)ds + \eta_n(t) + \varepsilon_n(t)$$

To take into account the effect of the estimation of the v_k, we will use the decomposition

$$\psi_{ij}v_i(t)v_j(s) = \hat{c}_i\psi_{ij}\hat{c}_j(\hat{c}_iv_i(t))(\hat{c}_jv_j(s))$$

$$= \hat{c}_i\psi_{ij}\hat{c}_j\hat{v}_i(t)\hat{v}_j(s)$$

$$+\hat{c}_i\psi_{ij}\hat{c}_j[\hat{c}_iv_i(t) - \hat{v}_i(t)]\hat{c}_jv_j(s) + \hat{c}_i\psi_{ij}\hat{c}_j\hat{v}_i(t)[\hat{c}_jv_j(s) - \hat{v}_j(s)]$$

which allows us to rewrite (1.1) as

$$Y_n(t) = \int \left(\sum_{i,j=1}^{p} \hat{c}_i\psi_{ij}\hat{c}_j\hat{v}_i(t)\hat{v}_j(s)\right) X_n(s)ds + \Delta_n(t).$$

■

To state Proposition 8.1, we introduce the vectors

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

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

$$\hat{\mathbf{\Delta}}_n = [\hat{\Delta}_{n1}, \hat{\Delta}_{n2}, \ldots, \hat{\Delta}_{np}]^T, \quad \hat{\Delta}_{nk} = \langle \Delta_n, \hat{v}_k\rangle.$$

Projecting relation (1.1) onto $\hat{v}_k$, we obtain by Lemma 8.1,

$$\langle Y_n, \hat{v}_k\rangle = \sum_{j=1}^{p} \hat{c}_k\psi_{kj}\hat{c}_j \langle X_n, \hat{v}_j\rangle + \langle \Delta_n, \hat{v}_k\rangle, \quad 1 \le k \le p,$$

from which the following proposition follows.

PROPOSITION 8.1 *If relation* (1.1) *holds with a Hilbert–Schmidt kernel* $\psi(\cdot,\cdot)$*, then*

$$\hat{\mathbf{Y}}_n = \widetilde{\mathbf{\Psi}}_p\hat{\mathbf{X}}_n + \hat{\mathbf{\Delta}}_n, \qquad n = 1, 2, \ldots N,$$

where $\widetilde{\mathbf{\Psi}}_p$ *is the* $p \times p$ *matrix with entries* $\hat{c}_k\psi_{kj}\hat{c}_j$, $k, j = 1, 2, \ldots p$.

To find the asymptotic distribution of the matrices $\mathbf{V}_h$, we establish several lemmas. Each of them removes terms which are asymptotically negligible, and in the process the leading terms are identified. Our first lemma shows that, asymptotically, in the definition of $\mathbf{V}_h$, the residuals

$$\mathbf{R}_n = \hat{\mathbf{Y}}_n - \widetilde{\mathbf{Y}}_n^{\wedge} = (\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\hat{\mathbf{X}}_n + \hat{\mathbf{\Delta}}_n. \tag{8.2}$$

can be replaced by the "errors" $\hat{\mathbf{\Delta}}_n$. The essential element of the proof is the relation $\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge} = O_P(N^{-1/2})$ stated in Proposition 3.1.

LEMMA 8.2 *Suppose Assumptions 2.1 and 3.1 and condition* (2.1) *hold. Then, for any fixed* $h > 0$,

$$\left|\left|\mathbf{V}_h - N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{\Delta}}_n\hat{\mathbf{\Delta}}_{n+h}^T\right|\right| = O_P(N^{-1}).$$

PROOF OF LEMMA 8.2. By (8.2) and (4.3),

$$\mathbf{V}_h = N^{-1}\sum_{n=1}^{N-h}[(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\hat{\mathbf{X}}_n + \hat{\mathbf{\Delta}}_n][(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\hat{\mathbf{X}}_{n+h} + \hat{\mathbf{\Delta}}_{n+h}]^T.$$

Denoting, $\hat{\mathbf{C}}_h = N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{X}}_n\hat{\mathbf{X}}_{n+h}^T$, we thus obtain

$$\mathbf{V}_h = (\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\hat{\mathbf{C}}_h(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})^T + (\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{X}}_n\hat{\mathbf{\Delta}}_{n+h}^T$$

$$+N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{\Delta}}_n\hat{\mathbf{X}}_{n+h}^T(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})^T + N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{\Delta}}_n\hat{\mathbf{\Delta}}_{n+h}^T.$$

By the CLT for h–dependent vectors, $\hat{\mathbf{C}}_h = O_P(1)$, so the first term satisfies

$$(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\hat{\mathbf{C}}_h(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})^T = O_P(N^{-1/2}N^{-1/2}) = O_P(N^{-1}).$$

To deal with the remaining three terms, we use the decomposition of Lemma 8.1. It is enough to bound the coordinates of each of the resulting terms. Since $\Delta_n = \varepsilon_n + \eta_{n1} + \eta_{n2} + \gamma_{n1} + \gamma_{n2}$, we need to establish bounds for $2 \times 5 = 10$ terms, but these bounds fall only to a few categories, so we will only deal with some typical cases.

Starting with the decomposition of $\hat{\mathbf{X}}_n\hat{\mathbf{\Delta}}_{n+h}^T$, observe that

$$N^{-1/2}\sum_{n=1}^{N-h}\langle X_n, \hat{v}_i\rangle\langle\varepsilon_{n+h}, \hat{v}_j\rangle = \iint\left(N^{-1/2}\sum_{n=1}^{N-h}X_n(t)\varepsilon_{n+h}(s)\right)\hat{v}_i(t)\hat{v}_j(s)dtds.$$

The terms $X_n(t)\varepsilon_{n+h}(s)$ are iid elements of the Hilbert space $L^2([0,1]\times[0,1])$, so by the CLT in a Hilbert space, see e.g. Section 2.3 of Bosq (2000),

$$\iint \left(N^{-1/2} \sum_{n=1}^{N-h} X_n(t)\varepsilon_{n+h}(s)dtds \right)^2 = O_P(1).$$

Since the $\hat{v}_j$ have unit norm, $\iint (\hat{v}_i(t)\hat{v}_j(s))^2 dtds = 1$. It therefore follows from the Cauchy–Schwarz inequality that

$$\sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \varepsilon_{n+h}, \hat{v}_j \rangle = O_P(N^{1/2}).$$

Thus, the ε_n contribute to $(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge}) N^{-1} \sum_{n=1}^{N-h} \hat{\mathbf{X}}_n \hat{\mathbf{\Delta}}_{n+h}^T$ a term of the order $O_P(N^{-1/2}N^{-1}N^{1/2}) = O_P(N^{-1})$, as required.

We now turn to the contribution of the $\eta_{n,1}$. As above, we have

$$N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \eta_{n+h,1}, \hat{v}_j \rangle = \iint \left(N^{-1/2} \sum_{n=1}^{N-h} X_n(t)\eta_{n+h,1}(s) \right) \hat{v}_i(t)\hat{v}_j(s)dtds$$

$$= \iint \left(N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \int \left(\sum_{k=p+1}^{\infty} \sum_{\ell=1}^{\infty} \psi_{k\ell} v_k(s) v_\ell(u) \right) X_{n+h}(u)du \right) \hat{v}_i(t)\hat{v}_j(s)dtds$$

$$= \int \left[\iint N_h(t,u) R_p(t,u) dtdu \right] v_k(s)\hat{v}_j(s)ds,$$

where

$$N_h(t,u) = N^{-1/2} \sum_{n=1}^{N-h} X_n(t) X_{n+h}(u)$$

and

$$R_p(t,u) = \sum_{\ell=1}^{\infty} \sum_{k=p+1}^{\infty} \psi_{k\ell} v_\ell(u) \hat{v}_k(t).$$

By the CLT for m–dependent elements in a Hilbert space, (follows e.g. from Theorem 2.17 of Bosq (2000)), $N_h(\cdot,\cdot)$ is $O_P(1)$ in $L^2([0,1]\times[0,1])$, so

$$\iint N_h^2(t,u)dtdu = O_P(1).$$

A direct verification using Assumption 3.1 shows that also

$$\iint R_p^2(t,u)dtdu = O_P(1).$$

Thus, by the Cauchy–Schwarz inequality, we obtain that

$$\sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \eta_{n+h,1}, \hat{v}_j \rangle = O_P(N^{1/2}),$$

and this again implies that the η_{n1} make a contribution of the same order as the ε_n. The same argument applies to the η_{n2}.

We now turn to the contribution of the γ_{n1}, the same argument applies to the γ_{n2}. Observe that, similarly as for the η_{n1},

$$N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle = \iint \left(N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \gamma_{n+h,1}(s) \right) \hat{v}_i(t) \hat{v}_j(s) dt ds \tag{8.3}$$

$$= \int \left[\iint N_h(t,u) \sum_{k,\ell=1}^{p} \hat{c}_k \psi_{k\ell} v_\ell(u) \hat{v}_i(t) dt du \right] [\hat{c}_k v_k(s) - \hat{v}_k(s)] \hat{v}_j(s) ds$$

Clearly,

$$\iint \left(\sum_{k,\ell=1}^{p} \hat{c}_k \psi_{k\ell} v_\ell(u) \hat{v}_i(t) \right)^2 dt du = O_P(1),$$

By Theorem 2.1,

$$\left\{ \int [\hat{c}_k v_k(s) - \hat{v}_k(s)]^2 ds \right\}^{1/2} = O_P(N^{-1/2}). \tag{8.4}$$

We thus obtain

$$\sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle = O_P(1), \tag{8.5}$$

so the contribution of γ_n is smaller than that of ε_n and η_n.

To summarize, we have proven that

$$(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge}) N^{-1} \sum_{n=1}^{N-h} \hat{\mathbf{X}}_n \hat{\mathbf{\Delta}}_{n+h}^T = O_P(N^{-1}).$$

The term $N^{-1} \sum_{n=1}^{N-h} \hat{\mathbf{\Delta}}_n \hat{\mathbf{X}}_{n+h}^T (\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})^T$ can be dealt with in a fully analogous way. ∎

By Lemma 8.1, the errors $\hat{\mathbf{\Delta}}_n$ can be decomposed as follows

$$\hat{\mathbf{\Delta}}_n = \hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n + \hat{\boldsymbol{\gamma}}_n,$$

with the coordinates obtained by projecting the functions $\varepsilon_n, \eta_n, \gamma_n$ onto the EFPC's $\hat{v}_j$. For example,

$$\hat{\boldsymbol{\eta}}_n = [\langle \eta_n, \hat{v}_1 \rangle, \langle \eta_n, \hat{v}_2 \rangle, \ldots, \langle \eta_n, \hat{v}_p \rangle]^T.$$

Lemma 8.3 shows that the vectors $\hat{\boldsymbol{\gamma}}_n$ do not contribute to the asymptotic distribution of the $\mathbf{V}_h$. This is essentially due to the fact that by Theorem 2.1, the difference between $\hat{v}_j$ and $\hat{c}_j v_j$ is of the order $O_P(N^{-1/2})$. For the same reason, in the definition of $\hat{\boldsymbol{e}}_n$ and $\hat{\boldsymbol{\eta}}_n$, the $\hat{v}_j$ can be replaced by the $\hat{c}_j v_j$, as stated in Lemma 8.4. Lemma 8.4 can be proven in a similar way as Lemma 8.3, so we present only the proof of Lemma 8.3.

LEMMA 8.3 *Suppose Assumptions 2.1 and 3.1 and condition* (2.1) *hold. Then, for any fixed* $h > 0$,

$$\left|\left| \mathbf{V}_h - N^{-1} \sum_{n=1}^{N-h} [\hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n][\hat{\boldsymbol{e}}_{n+h} + \hat{\boldsymbol{\eta}}_{n+h}]^T \right|\right| = O_P(N^{-1}).$$

LEMMA 8.4 *Suppose Assumptions 2.1 and 3.1 and condition* (2.1) *hold. Then, for any fixed* $h > 0$,

$$\left|\left| \mathbf{V}_h - N^{-1} \sum_{n=1}^{N-h} [\tilde{\boldsymbol{e}}_n + \tilde{\boldsymbol{\eta}}_n][\tilde{\boldsymbol{e}}_{n+h} + \tilde{\boldsymbol{\eta}}_{n+h}]^T \right|\right| = O_P(N^{-1}),$$

where

$$\tilde{\boldsymbol{e}}_n = [\hat{c}_1 \langle \varepsilon_n, v_1 \rangle, \hat{c}_2 \langle \varepsilon_n, v_2 \rangle, \ldots, \hat{c}_p \langle \varepsilon_n, v_p \rangle]^T$$

and

$$\begin{aligned} \tilde{\boldsymbol{\eta}}_n &= [\hat{c}_1 \langle \eta_n, v_1 \rangle, \hat{c}_2 \langle \eta_n, v_2 \rangle, \ldots, \hat{c}_p \langle \eta_n, v_p \rangle]^T \\ &= [\hat{c}_1 \langle \eta_{n2}, v_1 \rangle, \hat{c}_2 \langle \eta_{n2}, v_2 \rangle, \ldots, \hat{c}_p \langle \eta_{n2}, v_p \rangle]^T. \end{aligned}$$

PROOF OF LEMMA 8.3. In light of Lemma 8.2, we must show that the norm of difference between

$$N^{-1} \sum_{n=1}^{N-h} [\hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n][\hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n]^T$$

and

$$N^{-1} \sum_{n=1}^{N-h} [\hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n + \hat{\boldsymbol{\gamma}}_n][\hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n + \hat{\boldsymbol{\gamma}}_n]^T$$

is $O_P(N^{-1})$.

Writing $\hat{\boldsymbol{\eta}}_n = \hat{\boldsymbol{\eta}}_{n1} + \hat{\boldsymbol{\eta}}_{n2}$ and $\hat{\boldsymbol{\gamma}}_n = \hat{\boldsymbol{\gamma}}_{n1} + \hat{\boldsymbol{\gamma}}_{n2}$, we see that this difference consists of 20 terms which involve multiplication by $\hat{\boldsymbol{\gamma}}_{n1}$ or $\hat{\boldsymbol{\gamma}}_{n2}$. For example, analogously to (8.3), the term involving ε_n and and $\gamma_{n+h,1}$ has coordinates

$$N^{-1} \sum_{n=1}^{N-h} \langle \varepsilon_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle$$

$$= N^{-1/2} \int \left[\iint N_{\varepsilon,h}(t,u) \sum_{k,\ell=1}^{p} \hat{c}_k \psi_{k\ell} v_\ell(u) \hat{v}_i(t) dt du \right] [\hat{c}_k v_k(s) - \hat{v}_k(s)] \hat{v}_j(s) ds,$$

where

$$N_{\varepsilon,h}(t,u) = N^{-1/2} \sum_{n=1}^{N-h} \varepsilon_n(t) X_{n+h}(u).$$

By the argument leading to (8.5) (in particular by (8.4)),

$$N^{-1} \sum_{n=1}^{N-h} \langle \varepsilon_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle = O_P(N^{-1}).$$

The other terms can be bounded using similar arguments. The key point is that by (8.4), all these terms are $N^{1/2}$ times smaller than the other terms appearing in the decomposition of $N^{-1} \sum_{n=1}^{N-h} \hat{\mathbf{\Delta}}_n \hat{\mathbf{\Delta}}_n^T$.

■

No more terms can be dropped. The asymptotic approximation to $\mathbf{V}_h$ thus involves linear functionals of the following processes.

$$R_{N,h}^{(1)} = N^{-1/2} \sum_{n=1}^{N} \varepsilon_n(t) \varepsilon_{n+h}(s),$$

$$R_{N,h}^{(2)} = N^{-1/2} \sum_{n=1}^{N} \varepsilon_n(t) X_{n+h}(s),$$

$$R_{N,h}^{(3)} = N^{-1/2} \sum_{n=1}^{N} \varepsilon_{n+h}(t) X_n(s),$$

$$R_{N,h}^{(4)} = N^{-1/2} \sum_{n=1}^{N} X_n(t) X_{n+h}(s).$$

Lemma 8.5, which follows directly for the CLT in the space $L^2([0,1] \times [0,1])$ and the calculation of the covariances, summarizes the asymptotic behavior of the processes $R_{N,h}^{(i)}$.

LEMMA 8.5 *Suppose Assumptions 2.1 and 3.1 and condition* (2.1) *hold. Then*

$$\left\{ R_{N,h}^{(i)},\ 1 \le i \le 4,\ 1 \le h \le H \right\} \xrightarrow{d} \left\{ \Gamma_h^{(i)},\ 1 \le i \le 4,\ 1 \le h \le H \right\},$$

where the $\Gamma_h^{(i)}$ *are* $L^2([0,1] \times [0,1])$*–valued jointly Gaussian process such that the processes* $\left\{ \Gamma_h^{(i)},\ 1 \le i \le 4 \right\}$ *are independent and identically distributed.*

According to Lemmas 8.4 and 8.5, if

$$\hat{c}_1 = \hat{c}_2 = \ldots = \hat{c}_p = 1, \tag{8.6}$$

then

$$N^{1/2} \{\mathbf{V}_h,\ 1 \leq h \leq H\} \xrightarrow{d} \{\mathbf{T}_h,\ 1 \leq h \leq H\},$$

where the $\mathbf{T}_h,\ 1 \leq h \leq H$, are independent identically distributed normal random matrices. This is because the limit distribution of the $\mathbf{V}_h$ is determined by the random processes $R^{(i)}_{N,h},\ 1 \leq i \leq 4,\ 1 \leq h \leq H$ which are uncorrelated for every fixed N. Since their joint limit is multivariate normal, the asymptotic independence of the $R^{(i)}_{N,h},\ 1 \leq i \leq 4,\ 1 \leq h \leq H$ follows. This yields the asymptotic independence of $\mathbf{V}_1, \ldots \mathbf{V}_H$. Their asymptotic covariances can be computed using Lemma 8.1. After lengthy but straightforward calculations, the following lemma is established

LEMMA 8.6 *Suppose Assumptions 2.1 and 3.1 and condition* (2.1) *hold. If* (8.6) *holds, then for any fixed* $h > 0$*,*

$$N\ \mathrm{Cov}(\mathbf{V}_h(k,\ell), \mathbf{V}_h(k',\ell')) \to a(k,\ell;k',\ell'),$$

where

$$a(k,\ell;k',\ell')$$
$$= r_2(k,k')r_2(\ell,\ell') + r_2(k,k')r_1(\ell,\ell') + r_2(\ell,\ell')r_1(k,k') + r_1(k,k')r_1(\ell,\ell'),$$

with

$$r_1(\ell,\ell') = \sum_{j=p+1}^{\infty} \lambda_j \psi_{\ell j}\psi_{\ell' j}$$

and

$$r_2(k,k') = \iint E[\varepsilon_1(t)\varepsilon_1(s)]v_k(t)v_{k'}(s)dtds.$$

While assumption (8.6) is needed to obtain the asymptotic distribution of the autocovariance matrices $\mathbf{V}_h$, we will now show that it is possible to construct a test statistic which does not require assumption (8.6). The arguments presented below use a heuristic derivation, and the approximate equalities are denoted with "≈". The arguments could be formalized as in the proofs of Lemmas 8.3 and 8.4, but the details are not presented to conserve space.

We estimate $\langle \varepsilon_n, v_k \rangle$ by

$$\begin{aligned} e_{nk}^{\wedge} &= \langle Y_n, \hat{v}_k \rangle - \sum_{j=1}^{p} \tilde{\psi}_{kj}^{\wedge} \langle X_n, \hat{v}_j \rangle \\ &\approx \hat{c}_k \langle Y_n, v_k \rangle - \sum_{j=1}^{p} \hat{c}_k \psi_{kj} \hat{c}_j \hat{c}_j \langle X_n, v_j \rangle \\ &= \hat{c}_k \left(\langle Y_n, v_k \rangle - \sum_{j=1}^{p} \psi_{kj} \langle X_n, v_j \rangle \right) \\ &= \hat{c}_k \left(\langle \varepsilon_n, v_k \rangle + \sum_{j=p+1}^{\infty} \psi_{kj} \langle X_n, v_j \rangle \right). \end{aligned}$$

By the strong law of large numbers

$$\begin{aligned} &\frac{1}{N} \sum_{n=1}^{N} \left(\langle \varepsilon_n, v_k \rangle + \sum_{j=p+1}^{\infty} \psi_{kj} \langle X_n, v_j \rangle \right) \left(\langle \varepsilon_n, v_{k'} \rangle + \sum_{j=p+1}^{\infty} \psi_{k'j} \langle X_n, v_j \rangle \right) \\ &\overset{a.s.}{\to} E \left[\left(\langle \varepsilon_n, v_k \rangle + \sum_{j=p+1}^{\infty} \psi_{kj} \langle X_n, v_j \rangle \right) \left(\langle \varepsilon_n, v_{k'} \rangle + \sum_{j=p+1}^{\infty} \psi_{k'j} \langle X_n, v_j \rangle \right) \right] \\ &= r_1(k, k') + r_2(k, k'). \end{aligned}$$

Therefore, defining,

$$\hat{a}(k, k', \ell, \ell') = \left(\frac{1}{N} \sum_{n=1}^{N} e_{nk}^{\wedge} e_{nk'}^{\wedge} \right) \left(\frac{1}{N} \sum_{n=1}^{N} e_{n\ell}^{\wedge} e_{n\ell'}^{\wedge} \right),$$

we see that

$$\hat{a}(k, k', \ell, \ell') \approx \hat{c}_k \hat{c}_{k'} \hat{c}_\ell \hat{c}_{\ell'} a(k, k', \ell, \ell'). \tag{8.7}$$

By Lemma 8.6, under (8.6), the asymptotic covariance matrix of $N^{1/2}\mathrm{vec}(\mathbf{V}_h)$ is a $p^2 \times p^2$ matrix

$$\mathbf{M} = [\ \mathbf{A}(i,j),\ 1 \le i, j \le p\ ],$$

where

$$\mathbf{A}(i,j) = [\ a(\ell, i, k, j),\ 1 \le \ell, k \le p\ ].$$

By (8.7), an estimator of $\mathbf{M}$ is

$$\widehat{\mathbf{M}} = \left[\ \widehat{\mathbf{M}}(i,j),\ 1 \le i, j \le p\ \right],$$

where

$$\widehat{\mathbf{M}}(i,j) = [\ \hat{a}(\ell, i, k, j),\ 1 \le \ell, k \le p\ ].$$

Direct verification shows that $\widehat{\mathbf{M}}$ can be written in the form (4.4), which is convenient for coding.

As seen from (8.7), it cannot be guaranteed that the matrix $\widehat{\mathbf{M}}$ will be close to the matrix $\mathbf{M}$ because of the unknown signs $\hat{c}_i$. However, as will be seen in the proof of Theorem 5.1, statistic (4.5) does not depend on these signs.

Proof of Theorem 5.1. By Lemmas 8.2 and 8.3,

$$\operatorname{vec}(\mathbf{V}_h) = \operatorname{vec}\left(N^{-1} \sum_{n=1}^{N-h} [\hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n][\hat{\boldsymbol{e}}_{n+h} + \hat{\boldsymbol{\eta}}_{n+h}]^T \right) + O_P(N^{-1}).$$

The arguments used in the proof of Lemma 8.2 show that

$$\operatorname{vec}\left(N^{-1} \sum_{n=1}^{N-h} [\hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n][\hat{\boldsymbol{e}}_{n+h} + \hat{\boldsymbol{\eta}}_{n+h}]^T \right)$$

$$= [\widehat{\mathbf{C}} \otimes \widehat{\mathbf{C}}] \operatorname{vec}\left(N^{-1} \sum_{n=1}^{N-h} [\boldsymbol{e}_n + \boldsymbol{\eta}_n][\boldsymbol{e}_{n+h} + \boldsymbol{\eta}_{n+h}]^T \right) + o_P(1),$$

where the matrix $\widehat{\mathbf{C}}$ is defined by (3.8), and where

$$\boldsymbol{e}_n = [\langle \varepsilon_n, v_1 \rangle, \langle \varepsilon_n, v_2 \rangle, \ldots, \langle \varepsilon_n, v_p \rangle]^T;$$

$$\boldsymbol{\eta}_n = [\langle \eta_n, v_1 \rangle, \langle \eta_n, v_2 \rangle, \ldots, \langle \eta_n, v_p \rangle]^T.$$

Similar arguments also show that

$$\widehat{\mathbf{M}} = [\widehat{\mathbf{C}} \otimes \widehat{\mathbf{C}}]\mathbf{M}[\widehat{\mathbf{C}} \otimes \widehat{\mathbf{C}}] + o_P(1).$$

Since $[\widehat{\mathbf{C}} \otimes \widehat{\mathbf{C}}]^T[\widehat{\mathbf{C}} \otimes \widehat{\mathbf{C}}]$ is the $p^2 \times p^2$ identity matrix, we obtain by Lemma 8.5 that

$$Q_N^{\wedge} = N \sum_{h=1}^{H} \left\{ \operatorname{vec}\left(N^{-1} \sum_{n=1}^{N-h} [\boldsymbol{e}_n + \boldsymbol{\eta}_n][\boldsymbol{e}_{n+h} + \boldsymbol{\eta}_{n+h}]^T \right) \right.$$

$$\left. \mathbf{M}^{-1} \left[\operatorname{vec}\left(N^{-1} \sum_{n=1}^{N-h} [\boldsymbol{e}_n + \boldsymbol{\eta}_n][\boldsymbol{e}_{n+h} + \boldsymbol{\eta}_{n+h}]^T \right) \right]^T \right\} + o_P(1).$$

In particular, we see that the asymptotic distribution of $Q_N^{\wedge}$ does not depend on the signs $\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_p$ (the same argument shows that $Q_N^{\wedge}$ itself does not depend on these signs), so we may assume that they are all equal to 1. The claim then follows form Lemmas 8.5 and 8.6.

∎

9 Proof of Theorem 5.2

We use expansions with respect to the v_i as well as to the u_j, so we replace (3.1) by

$$\psi(t,s) = \sum_{i,j=1}^{\infty} \lambda_i^{-1} \sigma_{ij} u_j(t) v_i(s) \tag{9.1}$$

which leads to

$$\zeta_{nj} = \sum_{i=1}^{\infty} \xi_{ni} \lambda_i^{-1} \sigma_{ij} + \varepsilon_{nj}, \quad 1 \le j \le q,\ 1 \le n \le N. \tag{9.2}$$

where

$$\zeta_{nj} = \langle u_j, Y_n \rangle\,,\ \xi_{ni} = \langle v_i, X_n \rangle\,,\ \varepsilon_{nj} = \langle u_j, \varepsilon_n \rangle\,,$$
$$\sigma_{ij} = E[\xi_{ni}\zeta_{nj}],\ \lambda_i = E\xi_{ni}^2.$$

Introducing

$$\mathbf{Z}_n = [Z_{n1}, Z_{n2}, \ldots, Z_{nq}]^T, \quad 1 \le n \le N,$$

$$Z_{nj} := \zeta_{nj} - \sum_{i=1}^{p} \xi_{ni} \lambda_i^{-1} \sigma_{ij}, \quad 1 \le j \le q, \tag{9.3}$$

the vector of nonobservable residuals, by (9.2), we have

$$Z_{nj} = \varepsilon_{nj} + r_{nj}, \quad r_{nj} = \sum_{i=p+1}^{\infty} \xi_{ni} \lambda_i^{-1} \sigma_{ij}, \quad 1 \le j \le q. \tag{9.4}$$

Define by $\mathbf{C}_h$ the $q \times q$ autocovariance matrix with entries

$$c_h(k,\ell) = \frac{1}{N} \sum_{n=1}^{N-h} Z_{nk} Z_{n+h,\ell}$$

Analogously to the statistics $\hat{Q}_N$ of Theorem 5.2, define

$$Q_N = N \sum_{h=1}^{H} \sum_{i,j=1}^{q} r_{f,h}(i,j) r_{b,h}(i,j),$$

Where $r_{f,h}(i,j)$ and $r_{b,h}(i,j)$ the (i,j) entries, respectively, of $\mathbf{C}_0^{-1}\mathbf{C}_h$ and $\mathbf{C}_h\mathbf{C}_0^{-1}$.

Our first result is the limit distribution of Q_N.

PROPOSITION 9.1 *Under the assumptions of Theorem 5.2, Q_N converges in distribution to the χ^2–distribution with q^2H degrees of freedom.*

PROOF: Since the vectors $\mathbf{Z}_n$ with coordinates (9.3) are independent and identically distributed with zero mean and finite fourth moment, the result follows from Theorem B.3 of Gabrys and Kokoszka (2007).

∎

We must now show that $\hat{Q}_N - Q_N \xrightarrow{P} 0$. To do this, we must use Theorem 2.1, which in turn requires that Q_N and $\hat{Q}_N$ be invariant to the signs of the EFPC's. This property is established in the following lemma.

LEMMA 9.1 *Set $c_i = \text{sign}(\langle v_i, \hat{v}_i \rangle)$ and $d_i = \text{sign}(\langle u_i, \hat{u}_i \rangle)$. (i) The value of Q_N does not change if each v_i is replaced by $c_i v_i$ and each u_i by $d_i u_i$. (ii) The value of $\hat{Q}_N$ does not change if each $\hat{v}_i$ is replaced by $c_i \hat{v}_i$ and each $\hat{u}_i$ by $d_i \hat{u}_i$.*

PROOF: We will proof statement (i), the arguments for statement (ii) is the same, "hats" have to be added to all formulas. Denote all quantities obtained by using $c_i v_i$ and $d_i u_i$ in place of v_i and u_i with a prime $'$. Then

$$Z'_{jn} = d_j \zeta_{nj} - \sum_{i=1}^{p} c_i \xi_{ni} \lambda_i^{-1} c_i \sigma_{ij} d_j = d_j Z_{nj}.$$

Therefore,

$$c'_h(k,\ell) = \frac{1}{N} \sum_{n=1}^{N-h} d_k Z_{nk} d_\ell Z_{n\ell} = d_k d_\ell c_h(k,\ell).$$

Denoting $\boldsymbol{D} = \text{diag}(d_1, d_2, \ldots, d_q)$, we thus have $\mathbf{C}'_h = \boldsymbol{D} C_h \boldsymbol{D}$. Direct verification shows that $Q_N = \text{tr}[\mathbf{C}_h^T \mathbf{C}_0^{-1} \mathbf{C}_h \mathbf{C}_0^{-1}]$. Consequently,

$$Q'_N = \text{tr}[\boldsymbol{D}\mathbf{C}_h^T \boldsymbol{D}\boldsymbol{D}^{-1}\mathbf{C}_0^{-1}\boldsymbol{D}^{-1}\boldsymbol{D}\mathbf{C}_h\boldsymbol{D}\boldsymbol{D}^{-1}\mathbf{C}_0^{-1}\boldsymbol{D}^{-1}] = \text{tr}[\boldsymbol{D}\mathbf{C}_h^T\mathbf{C}_0^{-1}\mathbf{C}_h\mathbf{C}_0^{-1}\boldsymbol{D}^{-1}].$$

Since $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, it follows that $Q'_N = Q_N$.

∎

Proposition 9.1 and Lemma 9.1 show that to verify the convergence $\hat{Q}_N - Q_N \xrightarrow{P} 0$, we may replace $\hat{v}_i$ by $c_i \hat{v}_i$ and $\hat{u}_i$ by $d_i \hat{u}_i$ in the definition of $\hat{Q}_N$ to ensure that the differences $c_i \hat{v}_i - v_i$ and $d_i \hat{u}_i - u_i$ are small. The residuals $\hat{Z}_{nj}$ can be expressed as follows

$$\hat{Z}_{nj} = d_j Z_{nj} + (\hat{Z}_{nj} - Z_{nj}) + D_n(p),$$

$$D_n(p) = \sum_{i=1}^{p} \left[\hat{\lambda}_i^{-1} \hat{\sigma}_{ij} \hat{\xi}_{ni} - \lambda_i^{-1} d_j c_i \sigma_{ij} c_i \xi_{ni} \right].$$

In the following, we use the notation

$$v_{iN} = c_i \hat{v}_i, \qquad u_{iN} = d_i \hat{u}_i$$

and replace $\hat{v}_i$ and $\hat{u}_i$, respectively, by v_{iN} and u_{iN} in all definitions.

The following Lemma forms part of the proof of Theorem 1 of Gabrys and Kokoszka (2007).

LEMMA 9.2 *Under the assumptions of Theorem 5.2,*

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\xi}_{nk}(\xi_{n\ell}-\hat{\xi}_{n\ell}) = O_P(N^{-1/2})$$

and for $h \geq 1$,

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N-h}\hat{\xi}_{nk}(\xi_{n+h,\ell}-\hat{\xi}_{n+h,\ell}) = O_P(N^{-1}).$$

Analogous statements hold for the scores of the Y_n.

We will also use the following bounds.

LEMMA 9.3 *Under the assumptions of Theorem 5.2,*

$$\hat{\lambda}_j - \lambda_j = O_P(N^{-1/2}), \qquad \hat{\sigma}_{ij} - c_i d_j \sigma_{ij} = O_P(N^{-1/2}).$$

PROOF: The relation $\hat{\lambda}_j - \lambda_j = O_P(N^{-1/2})$ follows from Theorem 2.1. To establish the second relation, we start with the decomposition

$$\hat{\sigma}_{ij} - c_i d_j \sigma_{ij} = \frac{1}{N}\sum_{n=1}^{N}(\hat{\xi}_{ni}\hat{\zeta}_{nj} - \xi_{ni}\zeta_{nj}) + \frac{1}{N}\sum_{n=1}^{N}(\xi_{ni}\zeta_{nj} - E[\xi_{ni}\zeta_{nj}]).$$

The second term is $O_P(N^{-1/2})$ by the central limit theorem. The first term is further decomposed as

$$\frac{1}{N}\sum_{n=1}^{N}(\hat{\xi}_{ni}\hat{\zeta}_{nj} - \xi_{ni}\zeta_{nj}) = \frac{1}{N}\sum_{n=1}^{N}\hat{\xi}_{ni}(\hat{\zeta}_{nj} - \zeta_{nj}) + \frac{1}{N}\sum_{n=1}^{N}\zeta_{nj}(\hat{\xi}_{ni} - \xi_{ni}).$$

We will show that $N^{-1/2}\sum_{n=1}^{N}\zeta_{nj}(\hat{\xi}_{ni} - \xi_{ni})$ is bounded in probability, the other term is dealt with in a similar way. Observe that

$$\begin{aligned} N^{-1/2}\sum_{n=1}^{N}\zeta_{nj}(\hat{\xi}_{ni} - \xi_{ni}) &= N^{-1/2}\sum_{n=1}^{N}\langle Y_n, u_j\rangle\langle X_n, v_{iN} - v_i\rangle \\ &= \left\langle N^{-1/2}\sum_{n=1}^{N}\langle Y_n, u_j\rangle X_n, v_{iN} - v_i\right\rangle. \end{aligned}$$

By the strong law of large numbers in a Hilbert space, the norm of $N^{-1/2}\sum_{n=1}^{N}\langle Y_n, u_j\rangle X_n$ is $O_P(N^{1/2})$, and by Theorem 2.1, $||v_{iN} - v_i|| = O_P(N^{-1/2})$.

∎

Now we are ready to prove Proposition 9.2 which completes the proof of Theorem 5.2.

PROPOSITION 9.2 *If the assumptions of Theorem 5.2 holds, then*

$$\hat{\mathbf{C}}_0 - \mathbf{C}_0 = O_P(N^{-1/2}) \tag{9.5}$$

and for $h \geq 1$,

$$\hat{\mathbf{C}}_h - \mathbf{C}_h = O_P(N^{-1}). \tag{9.6}$$

PROOF: We will use a modified definition of the autocovariances $\hat{c}_h(k,\ell)$, namely

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

the sample means $\hat{\mu}_Z(k)$ add terms of the order $O_P(N^{-1})$ to the $\hat{c}_h(k,\ell)$.

To prove relations (9.5) and (9.6), we decompose $\hat{\mathbf{C}}_h - \mathbf{C}_h$ into a number of terms, and use Lemmas 9.2 and 9.3 to show that these terms are of an appropriate order in probability. Observe that for $h \geq 0$,

$$\hat{\mathbf{C}}_h - \mathbf{C}_h = \frac{1}{N}\sum_{h=1}^{N-h} \hat{Z}_{nk}(\hat{Z}_{n+h,\ell} - Z_{n+h,\ell}) + \frac{1}{N}\sum_{h=1}^{N-h} Z_{n+h,\ell}(\hat{Z}_{nk} - Z_{nk}) =: M_1 + M_2.$$

In the following, we consider only the first term, M_1, the same tools apply to M_2. We decompose M_1 as

$$M_1 = M_{11} + M_{12} + M_{13} + M_{14},$$

where

$$M_{11} = \frac{1}{N}\sum_{n=1}^{N-h} \hat{\zeta}_{nk}(\hat{\zeta}_{n+h,\ell} - \zeta_{n+h,\ell});$$

$$M_{12} = \sum_{j=1}^{p} \frac{1}{N}\sum_{n=1}^{N-h} \zeta_{nk}(\xi_{n+h,j}\lambda_j^{-1}\sigma_{j\ell} - \hat{\xi}_{n+h,j}\hat{\lambda}_j^{-1}\hat{\sigma}_{j\ell}) =: \sum_{j=1}^{p} M_{12j};$$

$$M_{13} = -\sum_{i=1}^{p} \frac{1}{N}\sum_{n=1}^{N-h} \hat{\xi}_{ni}\hat{\lambda}_i^{-1}\hat{\sigma}_{ik}(\hat{\zeta}_{n+h,\ell} - \zeta_{n+h,\ell}) =: -\sum_{i=1}^{p} M_{13i};$$

$$M_{14} = -\sum_{j=1}^{p}\sum_{i=1}^{p} \frac{1}{N}\sum_{n=1}^{N-h} \hat{\xi}_{ni}\hat{\lambda}_i^{-1}\hat{\sigma}_{ik}(\xi_{n+h,j}\lambda_j^{-1}\sigma_{j\ell} - \hat{\xi}_{n+h,j}\hat{\lambda}_j^{-1}\hat{\sigma}_{j\ell}) =: -\sum_{i=1}^{p}\sum_{j=1}^{p} M_{14ij}.$$

The term M_{11} is of correct order by Lemma 9.2.

Each term M_{12j} can be decomposed as $M_{12j} = M_{12j1} + M_{12j2}$, where

$$M_{12j1} = \frac{1}{N}\sum_{n=1}^{N-h} \zeta_{kn}\xi_{n+h,j}(\lambda_j^{-1}\sigma_{j\ell} - \hat{\lambda}_j^{-1}\hat{\sigma}_{j\ell});$$

$$M_{12j2} = \frac{1}{N}\sum_{n=1}^{N-h} \zeta_{nk}\hat{\lambda}_j^{-1}\hat{\sigma}_{j\ell}(\xi_{n+h,j} - \hat{\xi}_{n+h,j}).$$

By Lemma 9.3, $M_{12j1} = O_P(N^{-1/2})N^{-1}\sum_{n=1}^{N-h}\zeta_{kn}\xi_{n+h,j}$. If $h = 0$, $N^{-1}\sum_{n=1}^{N}\zeta_{nk}\xi_{j,n} = O_P(1)$ by the law of large numbers. If $h \geq 1$, $N^{-1}\sum_{n=1}^{N-h}\zeta_{nk}\xi_{n+h,j} = O_P(N^{-1/2})$, by the central limit theorem. The term M_{12j2} has the same rates. If $h = 0$, by the law of large numbers in a Hilbert space,

$$\frac{1}{N}\sum_{n=1}^{N}\zeta_{nk}(\xi_{nj} - \hat{\xi}_{nj}) = \left\langle \frac{1}{N}\sum_{n=1}^{N}\langle u_k, Y_n\rangle X_n, v_j - v_{jN}\right\rangle = O_P(1)O_P(N^{-1/2}).$$

If $h \geq 1$, by the central limit theorem in a Hilbert space,

$$\frac{1}{N}\sum_{n=1}^{N-h}\zeta_{nk}(\xi_{n+h,j} - \hat{\xi}_{n+h,j}) = \left\langle \frac{1}{N}\sum_{n=1}^{N-h}\langle u_k, Y_n\rangle X_{n+h}, v_j - v_{jN}\right\rangle = O_P(N^{-1/2})O_P(N^{-1/2}).$$

We conclude that

$$M_{12} = O_P(N^{-1/2}), \text{ if } h = 0; \quad M_{12} = O_P(N^{-1}), \text{ if } h \geq 1.$$

The same technique shows that for $\alpha = 3$ and $\alpha = 4$

$$M_{1\alpha} = O_P(N^{-1/2}), \text{ if } h = 0; \quad M_{1\alpha} = O_P(N^{-1}), \text{ if } h \geq 1.$$

∎

10 Proofs of Theorems 5.3 and 5.4

Proof of Theorem 5.3: We closely follow the plan of the proof of Theorem 5.1. The decomposition in Lemma 8.1 and Proposition 8.1 clearly hold for dependent X_n.

To formulate our first lemma, we introduce the $p \times p$ matrix

$$\widehat{\mathbf{K}}_p = \left[\hat{c}_i\hat{c}_j\sum_{k=p+1}^{\infty}\iint \psi_{jk}v_k(u)e_h(t,u)v_j(t)dtdu,\ 1 \leq i,j \leq p\right].$$

Lemma 10.1 *Under the assumptions of Theorem 5.3, for any fixed $h > 0$,*

$$\left|\left|\mathbf{V}_h - \left[N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{\Delta}}_n\hat{\mathbf{\Delta}}_{n+h}^T + (\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\widehat{\mathbf{K}}_p + \widehat{\mathbf{K}}_p^T(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})^T\right]\right|\right| = O_P(N^{-1}).$$

PROOF OF LEMMA 10.1. By (8.2) and (4.3),

$$\mathbf{V}_h = N^{-1}\sum_{n=1}^{N-h}[(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\hat{\mathbf{X}}_n + \hat{\mathbf{\Delta}}_n][(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\hat{\mathbf{X}}_{n+h} + \hat{\mathbf{\Delta}}_{n+h}]^T.$$

Denoting, $\hat{\mathbf{C}}_h = N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{X}}_n\hat{\mathbf{X}}_{n+h}^T$, we thus obtain

$$\mathbf{V}_h = (\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\hat{\mathbf{C}}_h(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})^T + (\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{X}}_n\hat{\mathbf{\Delta}}_{n+h}^T$$

$$+N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{\Delta}}_n\hat{\mathbf{X}}_{n+h}^T(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})^T + N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{\Delta}}_n\hat{\mathbf{\Delta}}_{n+h}^T.$$

By the ergodic theorem , $\hat{\mathbf{C}}_h = O_P(1)$, so the first term satisfies

$$(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})\hat{\mathbf{C}}_h(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})^T = O_P(N^{-1/2}N^{-1/2}) = O_P(N^{-1}).$$

To deal with the remaining three terms, we use the decomposition of Lemma 8.1. Starting with the decomposition of $\hat{\mathbf{X}}_n\hat{\mathbf{\Delta}}_{n+h}^T$, observe that

$$N^{-1/2}\sum_{n=1}^{N-h}\langle X_n, \hat{v}_i\rangle\langle\varepsilon_{n+h}, \hat{v}_j\rangle = \iint\left(N^{-1/2}\sum_{n=1}^{N-h}X_n(t)\varepsilon_{n+h}(s)\right)\hat{v}_i(t)\hat{v}_j(s)dtds.$$

It is verified in Aue *et al.* (2010) that

$$\iint\left(N^{-1/2}\sum_{n=1}^{N-h}X_n(t)\varepsilon_{n+h}(s)dtds\right)^2 = O_P(1).$$

Since the $\hat{v}_j$ have unit norm, it follows from the Cauchy–Schwarz inequality that

$$\sum_{n=1}^{N-h}\langle X_n, \hat{v}_i\rangle\langle\varepsilon_{n+h}, \hat{v}_j\rangle = O_P(N^{1/2}).$$

Thus, the ε_n contribute to $(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})N^{-1}\sum_{n=1}^{N-h}\hat{\mathbf{X}}_n\hat{\mathbf{\Delta}}_{n+h}^T$ a term of the order $O_P(N^{-1/2}N^{-1}N^{1/2}) = O_P(N^{-1})$.

We now turn to the contribution of the $\eta_{n,1}$. As above, we have

$$N^{-1/2}\sum_{n=1}^{N-h}\langle X_n, \hat{v}_i\rangle\langle\eta_{n+h,1}, \hat{v}_j\rangle = \iint\left(N^{-1/2}\sum_{n=1}^{N-h}X_n(t)\eta_{n+h,1}(s)\right)\hat{v}_i(t)\hat{v}_j(s)dtds$$

$$= \iint\left(N^{-1/2}\sum_{n=1}^{N-h}X_n(t)\int\left(\sum_{k=p+1}^{\infty}\sum_{\ell=1}^{\infty}\psi_{k\ell}v_k(s)v_\ell(u)\right)X_{n+h}(u)du\right)\hat{v}_i(t)\hat{v}_j(s)dtds.$$

Setting

$$N_h(t,u) = N^{-1/2} \sum_{n=1}^{N-h} [X_n(t)X_{n+h}(u) - e_h(t,u)]$$

and

$$R_p(s,u) = \sum_{\ell=1}^{\infty} \sum_{k=p+1}^{\infty} \psi_{k\ell} v_\ell(u) \hat{v}_i(s),$$

we thus obtain

$$N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \eta_{n+h,1}, \hat{v}_j \rangle$$

$$= \iiint N_h(t,u)\hat{v}_j(s)R_p(s,u)\hat{v}_i(t)dtdsdu$$

$$+\frac{N-h}{N^{1/2}} \iiint e_h(t,u)\hat{v}_j(s)R_p(s,u)\hat{v}_i(t)dtdsdu.$$

By the Cauchy–Schwarz inequality, we have

$$\left| \iiint N_h(t,u)\hat{v}_j(s)R_p(s,u)\hat{v}_i(t)dtdsdu \right|$$

$$\leq \left(\iiint N_h^2(t,u)\hat{v}_j^2(s)dtdsdu \right)^{1/2} \left(\iiint R_p^2(s,u)\hat{v}_i^2(t)dtdsdu \right)^{1/2}.$$

Aue *et al.* (2010) verified that

$$\iint N_h^2(t,u)dtdu = O_P(1). \tag{10.1}$$

A direct verification using Assumption 3.1 shows that also

$$\iint R_p^2(t,u)dtdu = O_P(1).$$

Hence

$$\left| \iiint N_h(t,u)\hat{v}_j(s)R_p(s,u)\hat{v}_i(t)dtdsdu \right| = O_P(1).$$

Using (2.2), we conclude that

$$\iiint e_h(t,u)\hat{v}_j(s)R_p(s,u)\hat{v}_i(t)dtdsdu$$

$$= \iiint e_h(t,u)v_j(s)R_p(s,u)v_i(t)dtdsdu + O_P(N^{-1/2}).$$

Since $j \leq p$, $\int R_p(s,u)v_j(s)ds = 0$, leading to

$$N^{-1/2}\sum_{n=1}^{N-h}\langle X_n, \hat{v}_i\rangle\langle \eta_{n+h,1}, \hat{v}_j\rangle = O_P(N^{1/2}). \tag{10.2}$$

Repeating the arguments leading to (3), we obtain that

$$N^{-1/2}\sum_{n=1}^{N-h}\langle X_n, \hat{v}_i\rangle\langle \eta_{n+h,2}, \hat{v}_j\rangle$$

$$= (N-h)\hat{c}_i\hat{c}_j\iiint e_h(t,u)v_j(s)\sum_{\ell=1}^{p}\sum_{k=p+1}^{\infty}\psi_{\ell k}v_\ell(s)v_k(u)v_i(t)dsdtdu + O_P(N^{1/2}).$$

By orthogonality of the v_i,

$$\int\sum_{\ell=1}^{p}\sum_{k=p+1}^{\infty}\psi_{\ell k}v_\ell(s)v_k(u)v_j(s)ds = \sum_{k=p+1}^{\infty}\psi_{jk}v_k(u).$$

We now turn to the contribution of the γ_{n1}, the same argument applies to the γ_{n2}. Observe that, similarly as for the η_{n1},

$$N^{-1/2}\sum_{n=1}^{N-h}\langle X_n, \hat{v}_i\rangle\langle \gamma_{n+h,1}, \hat{v}_j\rangle = \iint\left(N^{-1/2}\sum_{n=1}^{N-h}X_n(t)\gamma_{n+h,1}(s)\right)\hat{v}_i(t)\hat{v}_j(s)dtds \tag{10.3}$$

$$= \int\left[\iint N_h(t,u)\sum_{k,\ell=1}^{p}\hat{c}_k\psi_{k\ell}v_\ell(u)\hat{v}_i(t)dtdu\right][\hat{c}_kv_k(s) - \hat{v}_k(s)]\hat{v}_j(s)ds$$

$$+\frac{N-h}{N^{1/2}}\int\left[\iint e_h(t,u)\sum_{k,\ell=1}^{p}\hat{c}_k\psi_{k\ell}v_\ell(u)\hat{v}_i(t)dtdu\right][\hat{c}_kv_k(s) - \hat{v}_k(s)]\hat{v}_j(s)ds.$$

Clearly,

$$\iint\left(\sum_{k,\ell=1}^{p}\hat{c}_k\psi_{k\ell}v_\ell(u)\hat{v}_i(t)\right)^2 dtdu = O_P(1).$$

By (2.2),

$$\left\{\int[\hat{c}_kv_k(s) - \hat{v}_k(s)]^2ds\right\}^{1/2} = O_P(N^{-1/2})$$

and

$$\iiint e_h(t,u)\sum_{k,\ell=1}^{p}\hat{c}_k\psi_{k\ell}v_\ell(u)\hat{v}_i(t)[\hat{c}_kv_k(s) - \hat{v}_k(s)]\hat{v}_j(s)dtdsdu = O_P(N^{-1/2}).$$

We thus obtain

$$\sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle = O_P(N^{1/2}), \tag{10.4}$$

with the same bound holding for $\gamma_{n+h,2}$ in place of $\gamma_{n+h,1}$.

To summarize, we have expanded, as required, the term $(\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge}) N^{-1} \sum_{n=1}^{N-h} \hat{\mathbf{X}}_n \hat{\mathbf{\Delta}}_{n+h}^T$. The term $N^{-1} \sum_{n=1}^{N-h} \hat{\mathbf{\Delta}}_n \hat{\mathbf{X}}_{n+h}^T (\widetilde{\mathbf{\Psi}}_p - \widetilde{\mathbf{\Psi}}_p^{\wedge})$ can be dealt with in a fully analogous way. ∎

By Lemma 8.1, the errors $\hat{\mathbf{\Delta}}_n$ can be decomposed as follows

$$\hat{\mathbf{\Delta}}_n = \hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n + \hat{\boldsymbol{\gamma}}_n,$$

with the coordinates obtained by projecting the functions $\varepsilon_n, \eta_n, \gamma_n$ onto the EFPC's $\hat{v}_j$. For example,

$$\hat{\boldsymbol{\eta}}_n = [\langle \eta_n, \hat{v}_1 \rangle, \langle \eta_n, \hat{v}_2 \rangle, \ldots, \langle \eta_n, \hat{v}_p \rangle]^T.$$

Lemma 10.2 shows that $\hat{\boldsymbol{\gamma}}_n$ contributes a drift term to the asymptotic distribution of $\mathbf{V}_h$. To formulate it, we introduce the a $p \times p$ matrix $\widehat{\mathbf{F}}_p$ with entries

$$\begin{aligned}
\hat{F}_p(i,j) \;=\; & \sum_{m,r=1}^{p} \sum_{\ell=p+1}^{\infty} \psi_{i\ell}\psi_{mr} \iint v_\ell(s) v_r(z) e_h(s,z) ds dz \int \hat{c}_m [\hat{c}_m v_m(u) - \hat{v}_m(u)] v_j(u) du \\
& + \sum_{m=1}^{p} \sum_{\ell=1}^{\infty} \psi_{i\ell}\psi_{mj}\hat{c}_m \iint v_\ell(s) e_h(s,z) [\hat{c}_j v_j(s) - \hat{v}_j(s)] ds dz \\
& + \sum_{m,r=1}^{p} \sum_{\ell=p+1}^{\infty} \psi_{j\ell}\psi_{mr} \iint v_\ell(s) v_r(z) e_h(s,z) ds dz \int \hat{c}_m [\hat{c}_m v_m(u) - \hat{v}_m(u)] v_i(u) du \\
& + \sum_{m=1}^{p} \sum_{\ell=1}^{\infty} \psi_{j\ell}\psi_{mi}\hat{c}_m \iint v_\ell(s) e_h(s,z) [\hat{c}_j v_j(s) - \hat{v}_j(s)] ds dz.
\end{aligned}$$

LEMMA 10.2 *Under the assumptions of Theorem 5.3, for any fixed $h > 0$,*

$$\left\| N^{-1} \sum_{n=1}^{N-h} \hat{\mathbf{\Delta}}_n \hat{\mathbf{\Delta}}_{n+h}^T - \left\{ N^{-1} \sum_{n=1}^{N-h} [\hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n][\hat{\boldsymbol{e}}_{n+h} + \hat{\boldsymbol{\eta}}_{n+h}]^T + \widehat{\mathbf{F}}_p \right\} \right\| = O_P(N^{-1}).$$

PROOF OF LEMMA 10.2. Following the proof of Lemma 10.1, one can verify that

$$\left\| \sum_{n=1}^{N-h} \hat{\boldsymbol{e}}_n \hat{\boldsymbol{\eta}}_{n+h}^T \right\| = O_P(1), \qquad \left\| \sum_{n=1}^{N-h} \hat{\boldsymbol{\eta}}_n \hat{\boldsymbol{e}}_{n+h}^T \right\| = O_P(1)$$

and

$$\left\| \sum_{n=1}^{N-h} \hat{\boldsymbol{\eta}}_n \hat{\boldsymbol{\gamma}}_{n+h}^T \right\| = O_P(1), \qquad \left\| \sum_{n=1}^{N-h} \hat{\boldsymbol{\gamma}}_n^T \hat{\boldsymbol{\eta}}_{n+h,1}^T \right\| = O_P(1).$$

Using (2.2), (10.1), and the orthonormality of the v_i, we get

$$\sum_{n=1}^{N-h} \langle \eta_{n,2}, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle$$

$$= \hat{c}_i \hat{c}_j N \sum_{m,r=1}^{p} \sum_{l=p+1}^{\infty} \psi_{i\ell} \psi_{mr} \iint v_\ell(s) v_r(z) e_h(s,z) ds dz \hat{c}_m \psi_{mr} [\hat{c}_m v_m(u) - \hat{v}_m(u)] v_j(u) du + O_P(1)$$

and

$$\sum_{n=1}^{N-h} \langle \eta_{n,2}, \hat{v}_i \rangle \langle \gamma_{n+h,2}, \hat{v}_j \rangle$$

$$= \hat{c}_i \hat{c}_j N \sum_{m=1}^{p} \sum_{l=p+1}^{\infty} \psi_{i\ell} \psi_{mj} \iint v_\ell(s) v_r(z) e_h(s,z) ds dz \hat{c}_m \psi_{mr} [\hat{c}_j v_j(s) - \hat{v}_j(s)] ds dz + O_P(1).$$

The remaining two terms, $\sum_{n=1}^{N-h} \langle \gamma_{n,\alpha}, \hat{v}_i \rangle \langle \eta_{n+h,2}, \hat{v}_j \rangle$, $\alpha = 1, 2$, can be handled in the same way.

∎

To formulate the next lemma, we introduce the matrix $\widehat{\mathbf{G}}_p$ whose (i,j) entry is

$$\begin{aligned} \hat{G}_p(i,j) &= \hat{c}_j \sum_{k=1}^{p} \sum_{\ell,r=1}^{p} \psi_{k\ell} \psi_{jr} \iiint v_k(t) [\hat{v}_i(t) - \hat{c}_i v_i(t)] v_\ell(s) e_h(s,z) v_r(z) dt ds dz \\ &\quad + \hat{c}_i \sum_{m=1}^{p} \sum_{\ell,r=1}^{p} \psi_{i\ell} \psi_{mr} \iiint v_\ell(s) e_h(s,z) v_m(u) [\hat{v}_j(u) - \hat{c}_j v_j(u)] v_r(z) du ds dz. \end{aligned}$$

Recall also that

$$\tilde{\boldsymbol{e}}_n = [\hat{c}_1 \langle \varepsilon_n, v_1 \rangle, \hat{c}_2 \langle \varepsilon_n, v_2 \rangle, \ldots, \hat{c}_p \langle \varepsilon_n, v_p \rangle]^T$$

and

$$\tilde{\boldsymbol{\eta}}_n = [\hat{c}_1 \langle \eta_{n,2}, v_1 \rangle, \hat{c}_2 \langle \eta_{n,2}, v_2 \rangle, \ldots, \hat{c}_p \langle \eta_{n,2}, v_p \rangle]^T.$$

LEMMA 10.3 *Under the assumptions of Theorem 5.3, for any fixed $h > 0$,*

$$\left\| N^{-1} \sum_{n=1}^{N-h} [\hat{\boldsymbol{e}}_n + \hat{\boldsymbol{\eta}}_n][\hat{\boldsymbol{e}}_{n+h} + \hat{\boldsymbol{\eta}}_{n+h}]^T - \left\{ N^{-1} \sum_{n=1}^{N-h} [\tilde{\boldsymbol{e}}_n + \tilde{\boldsymbol{\eta}}_n][\tilde{\boldsymbol{e}}_{n+h} + \tilde{\boldsymbol{\eta}}_{n+h}]^T + \widehat{\mathbf{G}}_p \right\} \right\| = O_P(N^{-1}).$$

Lemma 10.3 can be established along the lines of the proof of Lemma 10.2.

PROOF OF THEOREM 5.4: The proof of Theorem 5.3 shows that

$$\left\| \mathbf{V}_h - \left[\hat{c}_i \hat{c}_j \left(V_h^{(e)}(i,j) + D_h(i,j) \right), \ 1 \le i, j \le p \right] \right\| \xrightarrow{P} 0,$$

where

$$V_h^{(e)}(i,j) = E[\langle \varepsilon_0, v_i\rangle \langle \varepsilon_h, v_j\rangle].$$

Thus, but the definition of the statistic $Q_N^{\wedge}$, the consistency is established if for some $1 \le h \le H$,

$$[\mathrm{vec}(\mathbf{V}_h^{(e)}) + \boldsymbol{D}_h)]^T \{\widehat{\mathbf{M}}_0 \otimes \widehat{\mathbf{M}}_0\}^{-1} \mathrm{vec}(\mathbf{V}_h^{(e)}) + \boldsymbol{D}_h) \neq 0.$$

The matrix $\boldsymbol{D}_h$ converges to zero, as $p \to \infty$. It can be shown, as in the proof of Theorem 5.3, that by imposing the same dependence conditions on the ε_n as on the X_n, we have $\hat{c}_i \hat{c}_j \hat{M}_0(i,j) \xrightarrow{P} M_0(i,j)$, where $M_0(i,j)$ is the limit of the empirical covariances $N^{-1}\sum_{n=1}^N e_{ni}^{\wedge} e_{ni}^{\wedge}$, and so is positive definite.

Thus, the test is consistent if for some $1 \le h \le H$,

$$[\mathrm{vec}(\mathbf{V}_h^{(e)})]^T \{\widehat{\mathbf{M}}_0 \otimes \widehat{\mathbf{M}}_0\}^{-1} \mathrm{vec}(\mathbf{V}_h^{(e)}) \neq 0,$$

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

∎

Acknowledgements. 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.

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.

Andersen, T. G. and Bollerslev, T. (1997b). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, **2–3,** 115–158.

Aue, A., Hörmann, S., Horváth, L. and Hušková, M. (2010). Sequential stability test for functional linear models. Technical Report. University of California Davis.

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*, **00,** 0000–0000; Forthcoming.

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.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer, New York.

Bowerman, B. L. and O'Connell, R. T. (1990). *Linear Statistical Models: an Applied Approach*, second edn. Duxbury.

Cai, T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, **34,** 2159–2179.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, New Jersey.

Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003). Testing hypothesis in the functional linear model. *Scandinavian Journal of Statistics*, **30,** 241–255.

Chatfield, C. (1998). Durbin–Watson test. In *Encyclopedia of Biostatistics* (eds P. Armitage and T. Colton), volume 2, pp. 1252–1253. Wiley.

Chiou, J-M. and Müller, H-G. (2007). Diagnostics for functional regression via residual processes. *Computational Statistics and Data Analysis*, **15,** 4849–4863.

Chiou, J-M., Müller, H-G. and Wang, J-L. (2004). Functional response models. *Statistica Sinica*, **14,** 675–693.

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.

Cook, D. R. (1977). Detection of influential observations in linear regression. *Technometrics*, **19,** 15–18.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Inference in Regression.* Chpaman and Hall.

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.

Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika*, **37,** 409–428.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, **38,** 159–178.

Durbin, J. and Watson, G. S. (1971). Testing for serial correlation in least squares regression. III. *Biometrika*, **58,** 1–19.

Gabrys, R. and Kokoszka, P. (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association*, **102,** 1338–1348.

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.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components. *J. R. Statist. Soc* B, **68,** 109–126.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57,** number 2, 357–384.

Hörmann, S. (2008). Augmented GARCH sequences: Dependence structure and asymptotics. *Bernoulli*, **14,** 543–561.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, **00,** 0000–0000; Forthcoming.

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

Horváth, L., Kokoszka, P. and Maslova, I. (2009). Change point detection in the functional linear model. Technical Report. Utah State University.

Horváth, L., Kokoszka, P. and Reimherr, M. (2009). Two sample inference in functional linear models. *Canadian Journal of Statistics*, **37,** 571–591; Forthcoming.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparcity for principal components analysis in high dimensions. *Journal of the Americal Statistical Association*, **104,** 682–693.

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.

Li, Y. and Hsing, T. (2007). On rates of convergence in functional linear regression. *Journal of Multivariate Analysis*, **98,** 1782–1804.

Malfait, N. and Ramsay, J. O. (2003). The historical functional model. *Canadian Journal of Statistics*, **31,** 115–128.

Maslova, I., Kokoszka, P., Sojka, J. and Zhu, L. (2009). Statistical significance testing for the association of magnetometer records at high–, mid– and low latitudes during substorm days. *Planetary and Space Science*, **00,** 000–000; Forthcoming.

Müller, H-G. and Stadtmüller, U. (2005). Generalized functional linear models. *The Annals of Statistics*, **33,** 774–805.

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*, **000,** 000–000; Forthcoming.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer.

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.

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.

Reiss, P. T. and Ogden, R. T. (2009a). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society (B)*, **71,** 505–523.

Reiss, P. T. and Ogden, R. T. (2009b). Functional generalized linear models with images as predictors. *Biometrics*, **00,** 000–000.

Sacks, J. and Ylvisaker, D. (1966). Designs for regression problems with correlated errors. *The Annals of Mathematical Statistics*, **37,** 66–89.

Schmoyer, R. L. (1994). Permutations test for correlation in regression errors. *Journal of the American Statistical Association*, **89,** 1507–1516.

Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis*. Wiley, New York.

Thiel, H. (1965). The analysis of disturbances in regression analysis. *Journal of the American Statistical Association*, **60,** 1067–1079.

Thiel, H. and Nagar, A. L. (1961). Testing the independence of regression disturbances. *Journal of the American Statistical Association*, **57,** 793–806.

Tsay, R. S. (2005). *Analysis of Financial Time Series*. Wiley.

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.

Yao, F., Müller, H-G. and Wang., J-L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, **33,** 2873–2903.