# Portmanteau Test of Independence for Functional Observations

Robertas GABRYS and Piotr KOKOSZKA

In various fields, observations are curves over some natural time interval. These curves often arise from finely spaced measurements (e.g., in physical sciences and finance) or from smoothing unequally spaced sparse measurements. Recent years have seen the development of tools for analyzing such data in the growing field of functional data analysis. To validate the assumptions underlying these tools, it is important to verify that the functional observations form a simple random sample. If the curves form a (functional) time series, then model validation typically relies on checking whether model residuals are independent and identically distributed. We propose a test for independence and identical distribution of functional observations. To reduce dimension, curves are projected on the most important functional principal components, then a test statistic based on lagged cross-covariances of the resulting vectors is constructed. We show that this dimension-reduction step introduces asymptotically negligible terms; that is, the projections behave asymptotically as iid vector–valued observations. A complete asymptotic theory based on correlations of random matrices, functional principal component expansions, and Hilbert space techniques is developed. The test statistic has a chi-squared asymptotic null distribution and can be readily computed using the `R` package `fda`. The test has good empirical size and power, which in our simulations and examples is not affected by the choice of the functional basis. Its application is illustrated on two data sets: credit card sales activity and geomagnetic records.

KEY WORDS: Functional observations; Independence test.

## 1. INTRODUCTION

The last two decades have seen the emergence of new technology allowing the collection and storage of time series of finely sampled observations. Every single trade on a speculative asset is recorded and stored, and so, for example, minute-by-minute values of an asset are available, resulting in 390 observations in a 6-1/2–hour trading day rather than a single closing price. Similar examples abound. Data of this type can be conveniently viewed as functional observations; for example, we treat the curve built up of 390 minute-by-minute values of an asset as a single observation.

Most inferential tools of functional data analysis (FDA) (see Ramsay and Silverman 2005) rely on the assumption of iid functional observations. In designed experiments (see, e.g., Müller and Stadtmüller 2005), this assumption can be ensured, but in observational data derived from time series, it requires verification. In traditional (nonfunctional) time series analysis, tests of independence and residual–based diagnostic tests derived from these tests play fundamental roles. In this article we propose a simple portmanteau test of independence for functional observations. In addition to its direct applicability, this test, with the underlying theory and numerical work presented here, likely will form the basis for residual–based specification tests for various functional time series models. Although there are many such tests for real– and vector–valued observations (see, e.g., Hosking 1980; Li 1981), no methodology is yet available for *functional* data.

The idea behind our test is simple and intuitively appealing. The functional observations $X_n(t)$, $t \in [0, T)$, $n = 1, 2, \ldots, N$, are approximated by the first $p$ terms of the principal component expansion,

$$X_n(t) \approx \sum_{k=1}^{p} X_{kn} v_k(t), \qquad n = 1, 2, \ldots, N. \tag{1}$$

The $v_k(t)$'s are the principal modes of variation [principal components (PCs)], and the $X_{kn}$'s are the random weights (scores) corresponding to each individual curve. For the sake of an intuitive argument, viewing the $v_k(t)$'s as deterministic curves, the iid assumption for the curves $X_n(\cdot)$ implies this assumption for the random vectors $[X_{1n}, \ldots, X_{pn}]'$. The method proposed by Chitturi (1976) can be used to test this assumption. In reality, the $v_k(t)$'s must be replaced by estimated PCs. This transition is not trivial, because the estimated PCs depend on all observations. Moreover, the difference between the population and sample PCs is of order $N^{-1/2}$, and thus the limit distribution of some statistics may contain an extra term. Our test statistic is based on products of scores, and we show that the effect of estimating the PCs is asymptotically negligible. We also show by simulations that it has only a small effect on the finite-sample performance of the test.

An example of the data that motivated this research is shown in Figure 1. Approximately 100 terrestrial geomagnetic observatories form a network, INTERMAGNET, designed to monitor and understand the behavior of currents of charged particles flowing in the magnetosphere and ionosphere. Modern digital magnetometers record three components of the magnetic field at 5-second resolution, but the data available at INTERMAGNET's website (*http://www.intermagnet.org*) or CDs consist of 1-minute averages (1,440 data points per day per component per observatory). The recent availability of these large data sets may lead to new insights if appropriate statistical tools can be developed. In this respect, FDA stands out due to its unique strength: Physicists naturally view magnetogram records as curves reflecting a continuous change in the structure of the various magnetic fields. In Figure 1, the daily variation caused by the Earth's rotation is prominent. Note, however, that the periodic pattern is not strong, because the current system is often nonlinearly disturbed by solar energy flows. For this reason, tools of traditional time series analysis based on stationarity, seasonality, and polynomial trends are not suitable here.

Robertas Gabrys is a Doctoral Candidate in Statistics (E-mail: *gabrys.robertas@gmail.com*) and Piotr Kokoszka is Professor of Statistics (E-mail: *Piotr.Kokoszka@usu.edu*), Department of Mathematics and Statistics, Utah State University, Logan, UT 84322. This research was supported by National Science Foundation grant DMS-0413653. The authors thank the joint editor, W. W. Piegorsch, an associate editor, and two referees for their constructive suggestions that substantially improved the manuscript, and also thank Lie Zhu and Algirdas Laukaitis for supplying the magnetometer and credit card transactions data.


Journal of the American Statistical Association
December 2007, Vol. 102, No. 480, Theory and Methods
DOI 10.1198/016214507000001111

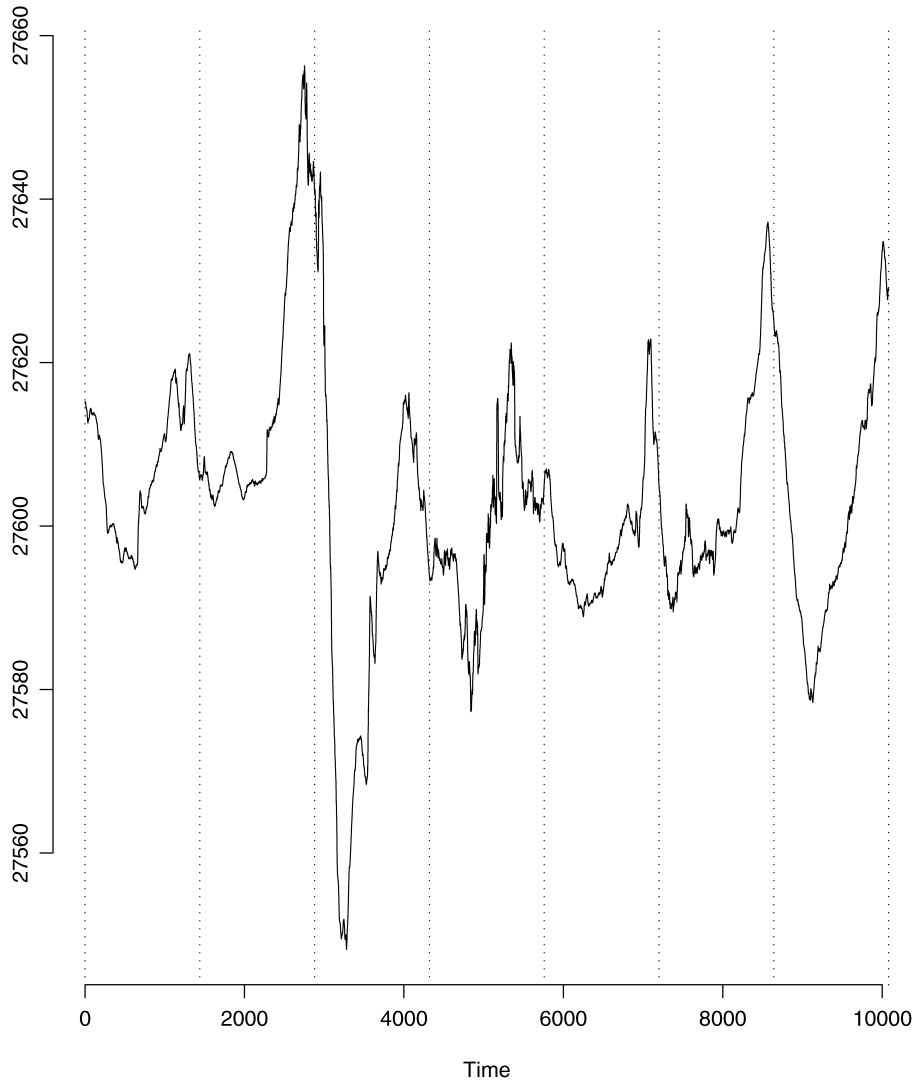


Figure 1. The horizontal component of the magnetic field measured in 1-minute resolution at Honolulu magnetic observatory from 1/1/2001 00:00 UT to 1/7/2001 24:00 UT.

The article is organized as follows. Section 2 formulates the test procedure along with mathematical assumptions and theorems establishing its asymptotic validity. The proofs of the theorems of Section 2 are presented in Section 5. Sections 3 and 4 are devoted, correspondingly, to the study of the finite-sample performance of the test and its application to two types of functional data: credit card sales and geomagnetic records. Appendix A contains some lemmas on the Hilbert space–valued random elements used in Section 5, which also may be useful in other, similar contexts. Appendix B develops the required theory for random matrices.

## 2. ASSUMPTIONS AND MAIN RESULTS

We first state the assumptions and introduce some notation. Our main result establishing limit null distribution is stated in Theorem 1. The consistency against the popular functional AR(1) model is established in Theorem 2.

We observe random functions $\{X_n(t), t \in [0,1), n = 1, 2, \ldots, N\}$ and want to test

$H_0$: the $X_n(\cdot)$ are independent and identically distributed versus $H_A$: $H_0$ does not hold.

We assume that the $X_n$'s are measurable elements of the space $L^2[0,1)$ in which the norm is defined by $\|X\|^2 = \int_0^1 X^2(t)\,dt$. For theoretical convenience, we assume that the $X_n$'s have mean 0. In Section 4 we explain how to center the data.

Our theory is valid under the assumption of finite fourth moments,

$$E\|X_n\|^4 = E\left[\int_0^1 X_n^2(t)\,dt\right]^2 < \infty. \tag{2}$$

If the $X_n$'s form a strictly stationary sequence (as is the case under $H_0$), then we denote by $X$ a random element with the distribution of the $X_n$'s and define the covariance operator

$$C(x) = E[\langle X, x\rangle X], \qquad x \in L^2[0,1).$$

The empirical covariance operator is defined by

$$C_N(x) = \frac{1}{N}\sum_{n=1}^{N}\langle X_n, x\rangle X_n, \qquad x \in L^2[0,1).$$

The eigenelements of $C$ are defined by

$$C(v_j) = \lambda_j v_j, \qquad j \geq 1.$$

The eigenfunctions $v_j$ form an orthonormal basis of $L^2[0, 1)$. We assume that $\lambda_1 \geq \lambda_2 \geq \lambda_3, \ldots$. The empirical eigenelements are defined by

$$C_N(v_{jN}) = \lambda_{jN} v_{jN}, \qquad j = 1, 2, \ldots, N,$$

where we assume that $\lambda_{1N} \geq \lambda_{2N} \geq \cdots \geq \lambda_{NN}$. Because the operators $C$ and $C_N$ are symmetric and nonnegative definite, the eigenvalues $\lambda_j$ and $\lambda_{jN}$ are nonnegative.

If the eigenspace corresponding to the eigenvalue $\lambda_k$ is one-dimensional, then formula (4.49) of Bosq (2000, p. 106) implies that

$$\limsup_{N\to\infty} N E\|v_k - v_{kN}\|^2 < \infty. \tag{3}$$

Note that $n$ before the expected value is missing in formula (4.49) of Bosq (2000); compare formulas (4.17) and (4.44) of that monograph.

To establish the null distribution of the test statistic $Q_N^F$, defined later, we merely require the following assumption:

*Assumption 1.* The functional observations $X_n$ are mean 0 in $L^2[0, 1)$, and (2) and (3) for $k \leq p$ hold.

The number $p$ (of PCs) appears in the statements of our main results.

We approximate the $X_n(t)$'s by

$$X_n^F(t) = \sum_{k=1}^{p} X_{kn}^F v_{kN}(t),$$

where

$$X_{kn}^F := \int_0^1 X_n^F(t) v_{kN}(t)\, dt = \int_0^1 X_n(t) v_{kN}(t)\, dt. \tag{4}$$

Thus the $X_n(t)$'s are the projections of the functional observations on the subspace spanned by the largest $p$ empirical PCs (see Ramsay and Silverman 2005, chap. 8).

We work with the random vectors

$$\mathbf{X}_n^F = [X_{1n}^F, X_{2n}^F, \ldots, X_{pn}^F]' \tag{5}$$

and analogously defined (unobservable) vectors

$$\mathbf{X}_n = [X_{1n}, X_{2n}, \ldots, X_{pn}]', \tag{6}$$

where

$$X_{kn} = \int_0^1 X_n(t) v_k(t)\, dt. \tag{7}$$

Under $H_0$, the $\mathbf{X}_n$'s are iid mean-0 random vectors in $R^p$ for which we denote

$$v(i, j) = E[X_{it} X_{jt}] \qquad \text{and} \qquad \mathbf{V} = [v(i, j)]_{i,j=1,\ldots,p}.$$

Thus the matrix $\mathbf{V}$ is the $p \times p$ covariance matrix. We let $\mathbf{C}_h$ denote the sample autocovariance matrix with entries

$$c_h(k, l) = \frac{1}{N} \sum_{t=1}^{N-h} X_{kt} X_{l,t+h}, \qquad 0 \leq h < N.$$

Let $r_{f,h}(i, j)$ and $r_{b,h}(i, j)$ denotes the $(i, j)$ entries of $\mathbf{C}_0^{-1}\mathbf{C}_h$ and $\mathbf{C}_h\mathbf{C}_0^{-1}$, and introduce the statistic

$$Q_N = N \sum_{h=1}^{H} \sum_{i,j=1}^{p} r_{f,h}(i, j) r_{b,h}(i, j). \tag{8}$$

Theorem B.3 shows that under $H_0$, $Q_N \xrightarrow{d} \chi^2_{p^2H}$.

Analogously to the way in which the statistic $Q_N$ (8) is constructed from the vectors $\mathbf{X}_n$, $n = 1, \ldots, N$, we construct the statistic $Q_N^F$ from the vectors $\mathbf{X}_n^F$, $n = 1, \ldots, N$.

The following theorem establishes the limit null distribution of the test statistic $Q_N^F$.

*Theorem 1.* Under $H_0$, if Assumption 1 holds, then $Q_N^F \xrightarrow{d} \chi^2_{p^2H}$.

This theorem is proven in Section 5.

Out of many possible directional alternatives, we focus on the ARH(1) (where $H$ stands for Hilbert space) model of Bosq (2000), which has been used in several interesting applications (see, e.g., Laukaitis and Račkauskas 2002; Fernández de Castro, Guillas, and Gonzáles Manteiga 2005; Kargin and Onatski 2007). It introduces serial correlation analogous to the usual AR(1) model.

Suppose then that

$$X_{n+1} = \Psi X_n + \varepsilon_n, \tag{9}$$

with iid mean-0 innovations $\varepsilon_n \in L^2[0, 1)$. We assume that the operator $\Psi$ is bounded, the solution to (9) is stationary, and Assumption 1 holds. These conditions are implied by very mild assumptions on $\Psi$ (see Bosq 2000, chaps. 3 and 4).

It is easy to check that the vectors (6) form a stationary VAR(1) process,

$$X_{k,n+1} = \sum_{i=1}^{p} \psi_{ki} X_{in} + \varepsilon_{k,n+1}, \qquad k = 1, 2, \ldots, p,$$

where $\psi_{ik} = \langle v_i, \Psi v_k \rangle$ and $\varepsilon_{k,n} = \langle \varepsilon_n, v_k \rangle$. In the vector form,

$$\mathbf{X}_{n+1} = \boldsymbol{\Psi}\mathbf{X}_n + \boldsymbol{\varepsilon}_{n+1}. \tag{10}$$

If the operator $\Psi$ is not 0, then there are $i, k \geq 1$ such that $\psi_{ik} \neq 0$.

The following theorem establishes the consistency against the ARH(1) model (9).

*Theorem 2.* Suppose that the functional observations $X_n$ follow a stationary solution to (9), Assumption 1 holds, and $p$ is so large that the $p \times p$ matrix $\boldsymbol{\Psi}$ in (10) is not 0. Then $Q_N^F \xrightarrow{P} \infty$.

This theorem is proven in Section 5. Other departures from the null can be treated in a similar manner. The idea is that under the null the sample autocovariances of the $X_{kn}^F$ at a positive lag converge in distribution when multiplied by $\sqrt{N}$, whereas under any reasonable alternative, these autocovariances tend in probability to some constants. This could be used to establish consistency against local alternatives, but this theoretical investigation is not pursued here.

## 3. FINITE–SAMPLE PERFORMANCE

In this section we investigate the finite-sample properties of the test using some generic models and sample sizes typical of applications discussed in Section 4. To investigate the empirical size, we generated independent trajectories of the standard Brownian motion (BM) on [0, 1] and the standard Brownian bridge (BB). We did this by transforming cumulative sums of independent normal variables computed on a grid of $m$ equispaced points in [0, 1]. We used values of $m$ ranging from 10 to 1,440, and found that the empirical size basically does not depend on $m$. (The tables of this section use $m = 100$.)

To compute the principal components $v_{kN}$ and the corresponding eigenvalues using the `R` package `fda`, the functional data must be represented (smoothed) using a specified number of functions from a basis. We worked with Fourier and B-spline functional bases and used 8, 16, and 80 basis functions. All results are based on 1,000 replications.

Table 1 shows empirical sizes for the BB and the Fourier basis for several values of the lag $H = 1, 3, 5$; the number of PCs $p = 16, 80$; and sample sizes $N = 50, 100, 300$. The standard errors in this table are between .5% and 1%. In most cases, the empirical sizes are within two standard errors of the nominal size, and the size improves somewhat with increasing $N$. The same is true for the BM and B-splines; no systematic dependence on the type of data or basis is seen, in accordance with the test's nonparametric nature. Of course, because the central limit theorem is used to establish the test's asymptotic validity, results are likely to be worse for nonnormal data, but a detailed empirical study is beyond the intended scope of this article.

In a power study, we focus on the ARH(1) model (9), which can be written more explicitly as

$$X_n(t) = \int_0^1 \psi(t,s) X_{n-1}(s)\, ds + \varepsilon_n(t), \qquad t \in [0,1], n = 1, 2, \ldots, N. \quad (11)$$

A sufficient condition for the assumptions of Theorem 2 to hold is

$$\|\Psi\|^2 = \int_0^1 \int_0^1 \psi^2(t,s)\, dt\, ds < 1. \quad (12)$$

Table 1. Empirical size (in percent) of the test using the Fourier basis

| | $p=3$ | | | $p=4$ | | | $p=5$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Lag | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| $N=50$ | | | | | | | | | |
| 1 | 7.7 | 2.5 | .6 | 7.4 | 2.8 | .3 | 7.9 | 3.5 | .4 |
| 3 | 6.8 | 2.5 | .3 | 6.7 | 3.3 | .6 | 4.9 | 2.0 | .3 |
| 5 | 4.9 | 2.0 | .0 | 3.6 | 1.4 | .2 | 4.0 | 1.7 | .2 |
| $N=100$ | | | | | | | | | |
| 1 | 9.0 | 5.1 | .4 | 8.9 | 3.9 | .6 | 10.0 | 3.9 | .8 |
| 3 | 8.1 | 3.5 | .6 | 8.3 | 4.0 | .9 | 7.5 | 3.2 | .4 |
| 5 | 8.8 | 3.6 | .6 | 6.6 | 2.7 | .3 | 6.7 | 2.4 | .3 |
| $N=300$ | | | | | | | | | |
| 1 | 9.8 | 4.6 | 1.2 | 9.4 | 4.0 | .9 | 10.1 | 4.7 | .6 |
| 3 | 9.3 | 4.8 | 1.0 | 9.1 | 4.7 | .9 | 10.0 | 5.4 | .8 |
| 5 | 7.2 | 3.7 | 1.0 | 8.2 | 3.8 | .7 | 10.6 | 5.5 | 1.2 |

NOTE: The simulated observations are BBs.

Table 2. Empirical power of the test using the Fourier basis

| | $p=3$ | | | $p=4$ | | | $p=5$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Lag | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| $N=50$ | | | | | | | | | |
| 1 | 44.7 | 33.8 | 17.7 | 41.9 | 29.4 | 12.6 | 38.5 | 26.1 | 9.2 |
| 3 | 35.2 | 27.0 | 13.3 | 34.0 | 24.7 | 10.8 | 33.2 | 21.6 | 8.7 |
| 5 | 26.7 | 20.0 | 11.0 | 24.4 | 15.8 | 8.1 | 21.5 | 14.3 | 6.0 |
| $N=100$ | | | | | | | | | |
| 1 | 71.2 | 64.2 | 51.4 | 74.4 | 66.5 | 48.1 | 77.7 | 68.0 | 46.1 |
| 3 | 67.9 | 61.0 | 44.9 | 67.5 | 58.6 | 42.8 | 68.4 | 56.9 | 38.1 |
| 5 | 62.3 | 54.6 | 38.6 | 59.0 | 49.9 | 32.3 | 55.1 | 45.5 | 27.9 |
| $N=300$ | | | | | | | | | |
| 1 | 98.7 | 98.2 | 96.7 | 99.2 | 98.9 | 97.2 | 99.8 | 99.5 | 98.5 |
| 3 | 97.6 | 97.1 | 95.5 | 99.0 | 98.4 | 96.8 | 99.2 | 98.3 | 96.6 |
| 5 | 96.8 | 95.9 | 92.8 | 98.1 | 97.0 | 93.8 | 98.4 | 97.3 | 94.4 |

NOTE: The observations follow the ARH(1) model (11) with Gaussian kernel with $\|\Psi\| = .5$ and iid standard BM innovations.

The norm in (12) is the Hilbert–Schmidt norm.

In our study, the innovations $\varepsilon_n$ in (11) are either standard BMs or BBs. We used two kernel functions, the Gaussian kernel,

$$\psi(t,s) = C \exp\left\{\frac{t^2+s^2}{2}\right\}, \qquad (t,s) \in [0,1]^2,$$

and the Wiener kernel,

$$\psi(t,s) = C \min(s,t), \qquad (t,s) \in [0,1]^2.$$

We chose the constants $C$ so that $\|\Psi\| = .3, .5, .7$. We used both Fourier and B-spline bases.

The power against this alternative is expected to increase rapidly with $N$, because the test statistic is proportional to $N$. This can be clearly seen in Table 2. The power also increases with $\|\Psi\|$; for $\|\Psi\| = .7$ and the Gaussian kernel, it is practically 100% for $N = 100$ and all choices of other parameters.

We also make two less trivial observations. The power is highest for lag $H = 1$. This is because for the ARH(1) process, the "correlation" between $X_n$ and $X_{n-1}$ is largest at this lag. By increasing the maximum lag $H$, the value of $Q_N^F$ generally increases, as does the critical value increases (the degrees of freedom increase by $p^2$ for a unit increase in $H$). The power also depends on how the action of the operator $\Psi$ is "aligned" with the eigenfunctions $v_k$. If the inner products $\langle v_i, \Psi v_k \rangle$ are large in absolute value, then the power is high. Thus, with all other parameters being the same, the power in Table 3 is greater than that in Table 2 because of the different covariance structures of the BB and BM. In all cases, the power for the Wiener kernel is slightly lower than that for the Gaussian kernel.

A full set of size and power tables is available on request. The empirical sizes for the two processes that we simulated are generally fairly close to nominal sizes and are not significantly affected by the choice of $H$ and $p$. Power against the ARH(1) model is very good if $H = 1$ is used.

## 4. APPLICATION TO CREDIT CARD TRANSACTIONS AND DIURNAL GEOMAGNETIC VARIATION

In this section we apply our test to two data sets that naturally lend themselves to FDA. The first data set, studied by Laukaitis

Table 3. Empirical power of the test using the Fourier basis

| | $p=3$ | | | $p=4$ | | | $p=5$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Lag | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| $N=50$ | | | | | | | | | |
| 1 | 98.3 | 97.0 | 92.1 | 98.4 | 96.4 | 87.6 | 99.1 | 97.3 | 87.6 |
| 3 | 95.2 | 90.3 | 77.4 | 92.1 | 86.2 | 69.6 | 89.9 | 85.1 | 63.2 |
| 5 | 86.9 | 80.2 | 61.7 | 78.5 | 71.7 | 51.4 | 75.2 | 63.9 | 40.4 |
| $N=100$ | | | | | | | | | |
| 1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3 | 100 | 100 | 99.9 | 100 | 100 | 99.7 | 100 | 99.9 | 99.8 |
| 5 | 99.9 | 99.3 | 98.7 | 99.9 | 99.8 | 98.6 | 99.7 | 99.5 | 97.8 |

NOTE: The observations follow the ARH(1) model (11) with Gaussian kernel with $\|\Psi\| = .5$ and iid standard BB innovations.

and Račkauskas (2002), consists of the number of transactions with credit cards issued by Vilnius Bank, Lithuania. The second is a daily geomagnetic variation; a similar data set was studied by Xu and Kamide (2004).

Suppose that $D_n(t_i)$ is the number of credit card transactions in day $n$, $n = 1, \ldots, 200$, (November 3, 2000–May 21, 2001) between times $t_{i-1}$ and $t_i$, where $t_i - t_{i-1} = 8$ minutes, $i = 1, \ldots, 128$. For our analysis, the transactions were normalized to have time stamps in interval $[0, 1]$, which corresponds to 1 day. To remove the weekly periodicity, we worked with the differences $X_n(t_i) = D_n(t_i) - D_{n-7}(t_i)$, $n = 1, 2, \ldots, 193$. Figure 2 displays the first 3 weeks of these data. A characteristic pattern of an AR(1) process with clusters of positive and negative observations can be clearly seen. Two consecutive days are shown in Figure 3(a) together with functional objects obtained by smoothing with 40 and 80 Fourier basis functions. As expected, the test rejects the null hypothesis at 1% level for both smooths, and all lag values $1 \le H \le 5$ and the number of PCs equal to 4, 5, 10, and 20.

We estimated the ARH(1) model (11) using the function `linmod` from the `R` package `fda` (see Malfait and Ramsay 2003; Ramsay and Silverman 2002, 2005). Table 4 displays the $p$ values that support this model choice (see Laukaitis and Račkauskas 2002). Note that starting with $p = 2$ or 3, the $p$ values increase and approach 100%. This is in agreement with the findings of Section 3 and is caused by the fact that the dependence is captured by only a few most important PCs, and increasing $p$ does not significantly change the sampling distribution of $Q_N^F$ but does shift the limiting distribution to the

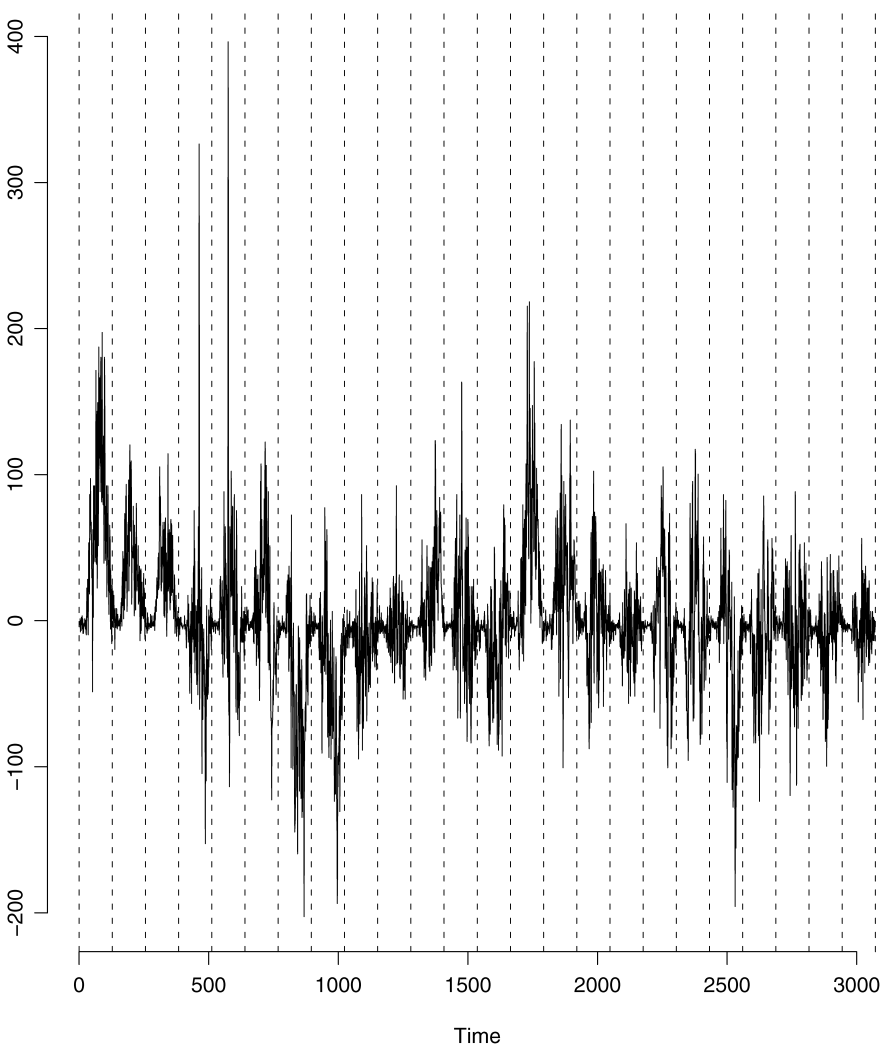


Figure 2. Three weeks of centered time series of $\{X_n(t_i)\}$ derived from the credit card transaction data. The vertical dotted lines separate days.

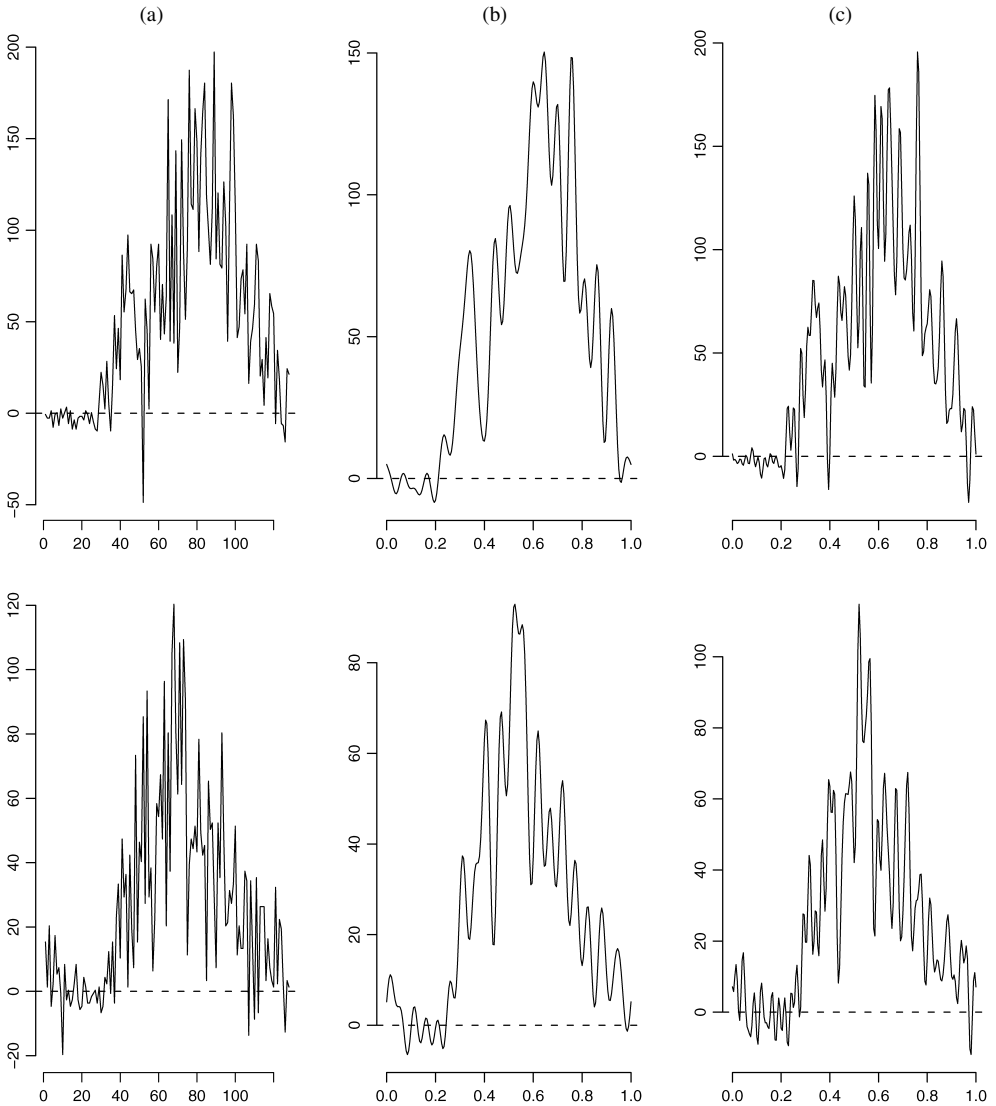


Figure 3. Two functional observations $X_n$ derived from the credit card transactions (a) together with smooths obtained by projection on 40 (b) and 80 (c) Fourier basis functions.

right. We note that, in analogy with well-known results for real-valued time series (see, e.g., Ljung and Box 1978 and references therein), the number of degrees of freedom in the asymptotic distribution of the statistic $Q_N^F$ computed from residuals of the ARH(1) model is likely to be less than $p^2H$. This question is a subject of ongoing research, but even with this caveat, Table 4 gives strong support to the ARH(1) model.

Table 4. $P$ values for the functional ARH(1) residuals of the credit card data $X_n$

| Lag, $H$ | $p=1$ | $p=2$ | $p=3$ | $p=4$ | $p=5$ | $p=6$ | $p=7$ |
|---|---|---|---|---|---|---|---|
| $BF=40$ | | | | | | | |
| 1 | 69.54 | 22.03 | 13.60 | 46.29 | 80.35 | 96.70 | 99.20 |
| 2 | 35.57 | 38.28 | 7.75 | 47.16 | 64.92 | 95.00 | 99.04 |
| 3 | 54.44 | 53.63 | 25.28 | 52.61 | 71.33 | 86.84 | 94.93 |
| $BF=80$ | | | | | | | |
| 1 | 57.42 | 18.35 | 53.30 | 89.90 | 88.33 | 95.40 | 99.19 |
| 2 | 35.97 | 23.25 | 23.83 | 45.07 | 55.79 | 46.39 | 70.65 |
| 3 | 36.16 | 36.02 | 26.79 | 30.21 | 56.81 | 34.51 | 47.00 |

We now turn to the ground-based magnetogram records, which reflect the variations of the currents flowing in the magnetosphere/ionosphere. These data are used to understand the structure of this important complex geosystem. Because here we present merely an illustration of our procedure, we focus only on the horizontal intensity measured at Honolulu in 2001. The horizontal ($H$) intensity is the component of the magnetic field tangent to the Earth's surface and pointing toward the magnetic north; its variation best reflects the changes in the large currents flowing in the magnetic equatorial plane. Figure 4(a) shows 2 weeks of these data. Xu and Kamide (2004) used the $H$ component measured at Beijing in 2001 to understand the statistical structure of the daily variation and associate it with known or conjectured currents. Following Xu and Kamide (2004), we subtracted the linear change over 1 day to obtain the curves like those shown in Figure 4(b). After centering over a period under study, we obtain the functional observations with

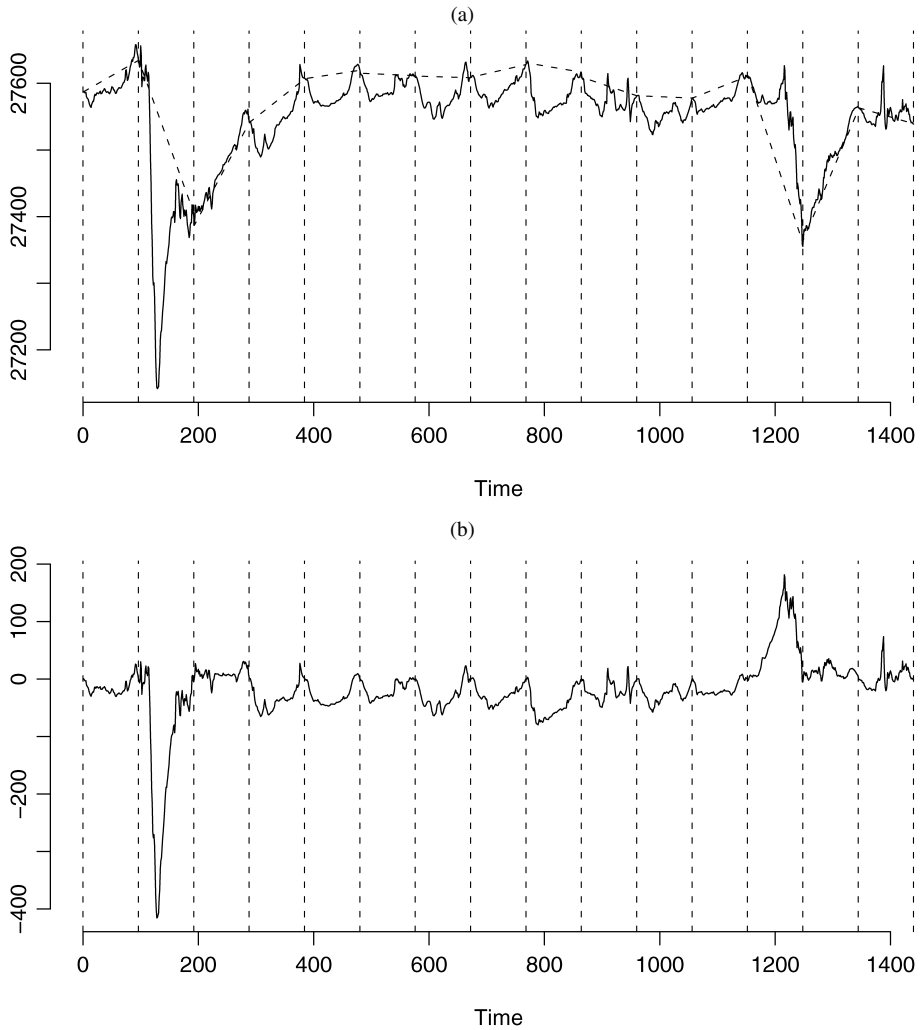


Figure 4. (a) Horizontal intensity (nT) measured at Honolulu between March 30, 2001 and April 13, 2001, with the straight lines connecting first and last measurements in each day. (b) The same after subtracting the lines.

which we work. The analysis was conducted using Fourier basis functions.

We note that the issue of separating the daily variation from larger disturbances caused by magnetic storms is a complex one and is the subject of ongoing geophysical research (see Jach, Kokoszka, Sojka, and Zhu 2006 for a recent contribution). For example, one can question whether what we see in the second day in Figure 4(b) is an unusually large daily variation or an unremoved signature of a magnetic storm. This article is not concerned with such issues, however.

Testing 1-year magnetometer data with lags $H = 1, 2, 3$ and different numbers of PCs, $p = 3, 4, 5$, yields $p$ values very close to 0. This indicates that while PC analysis, as advocated by Xu and Kamide (2004), may be a useful exploratory tool for studying daily variation over the whole year, we must be careful when using any inferential tools based on it, because these typically require a simple random sample (see, e.g., Seber 1984, sec. 5.2). We also applied the test to smaller subsets of data roughly corresponding to boreal spring and summer. The $p$ values reported in Table 5 show that the transformed data can, to a reasonable approximation be viewed as a functional simple random sample. The discrepancy in the test outcomes when applied to the whole year and to a single season is probably due to the annual change of the position of the Honolulu observatory relative to the sun, whose energy drives the convective currents mainly responsible for the daily variation.

The two examples discussed in this section show that our test can detect departures from the assumption of independence (credit card data) or from the assumption of identical distribution (magnetometer data), and confirm both assumptions when they are expected to hold. In our examples, the results of the test do not depend much on the choice of the functional basis.

## 5. PROOFS OF THEOREMS 1 AND 2

### Proof of Theorem 1

By Theorem B.3 in Appendix B, it is sufficient to show that $Q_N^F - Q_N \xrightarrow{P} 0$. By (8), this will follow if we show that for

Table 5. $p$ values for the magnetometer data split by season

| Lag | February–May | | June–September | |
|---|---|---|---|---|
| $H$ | $p = 4$ | $p = 5$ | $p = 4$ | $p = 5$ |
| 1 | 13.44 | 6.51 | 1.03 | 1.23 |
| 3 | 3.37 | 2.99 | 31.72 | 42.59 |

$h \geq 1$,

$$N^{1/2}(\mathbf{C}_h^F - \mathbf{C}_h) \xrightarrow{P} 0 \tag{13}$$

and $\mathbf{C}_0^F - \mathbf{C}_0 \xrightarrow{P} 0$. We will verify that (13) holds for all $h \geq 0$. Recall that

$$c_h(k,l) = \frac{1}{N}\sum_{n=1}^{N-h} X_{kn}X_{l,n+h} \qquad \text{and}$$

$$c_h^F(k,l) = \frac{1}{N}\sum_{n=1}^{N-h} X_{kn}^F X_{l,n+h}^F.$$

Therefore, $c_h^F(k,l) - c_h(k,l) = M_1 + M_2$, where

$$M_1 = \frac{1}{N}\sum_{n=1}^{N-h}(X_{kn} - X_{kn}^F)X_{l,n+h} \qquad \text{and}$$

$$M_2 = \frac{1}{N}\sum_{n=1}^{N-h} X_{kn}^F(X_{l,n+h} - X_{l,n+h}^F).$$

We will first show that $N^{1/2}M_1 \xrightarrow{P} 0$. Observe that

$$\begin{aligned} N^{1/2}M_1 &= N^{-1/2}\sum_{n=1}^{N-h}\langle X_n, v_k - v_{kN}\rangle\langle X_{n+h}, v_l\rangle \\ &= \left\langle N^{-1/2}\sum_{n=1}^{N-h}\langle X_{n+h}, v_l\rangle X_n, v_k - v_{kN}\right\rangle \\ &= \langle S_N, Y_N\rangle, \end{aligned}$$

where

$$S_N := N^{-1/2}\sum_{n=1}^{N-h}\langle X_{n+h}, v_l\rangle X_n \qquad \text{and} \qquad Y_N = v_k - v_{kN}.$$

Note that by (3),

$$\begin{aligned} E|\langle S_N, Y_N\rangle| &\leq E[\|S_N\|\,\|Y_N\|] \leq (E\|S_N\|^2)^{1/2}(E\|Y_N\|^2)^{1/2} \\ &= O\big(N^{-1/2}\big)(E\|S_N\|^2)^{1/2}. \end{aligned}$$

To show that $N^{1/2}M_1 \xrightarrow{P} 0$, it thus remains to verify that $E\|S_N\|^2$ is bounded. Note that

$$\begin{aligned} E\|S_N\|^2 &= N^{-1}E\left\|\sum_{n=1}^{N-h}\langle X_{n+h}, v_l\rangle X_n\right\|^2 \\ &= N^{-1}E\sum_{m,n=1}^{N-h}\langle X_{m+h}, v_l\rangle\langle X_{n+h}, v_l\rangle\langle X_m, X_n\rangle \\ &= N^{-1}\sum_{n=1}^{N-h}E[\langle X_{n+h}, v_l\rangle]^2 E\|X_n\|^2 \\ &\leq [E\|X_n\|^2]^2. \end{aligned}$$

To show that $N^{1/2}M_2 \xrightarrow{P} 0$, decompose $M_2$ as $M_2 = M_{21} + M_{22}$, where

$$M_{21} = \frac{1}{N}\sum_{n=1}^{N-h}\langle X_n, v_k\rangle\langle X_{n+h}, v_l - v_{lN}\rangle$$

and

$$M_{22} = \frac{1}{N}\sum_{n=1}^{N-h}\langle X_n, v_{kN} - v_k\rangle\langle X_{n+h}, v_l - v_{lN}\rangle.$$

By the argument developed for $M_1$, $N^{1/2}M_{21} \xrightarrow{P} 0$, so we must show that $N^{1/2}M_{22} \xrightarrow{P} 0$. This follows from Lemma A.2 in Appendix A.

## Proof of Theorem 2

We first state Lemmas 1 and 2, which form two critical building blocks of the proof.

*Lemma 1.* Suppose that the vectors $\mathbf{X}_n = [X_{1n}, X_{2n}, \ldots, X_{pn}]'$ follow a stationary vector AR(1) process $\mathbf{X}_{n+1} = \boldsymbol{\Psi}\mathbf{X}_n + \boldsymbol{\varepsilon}_{n+1}$. The errors $\boldsymbol{\varepsilon}_n$ are iid mean 0 with finite variance and $\boldsymbol{\varepsilon}_{n+1}$ is independent of $\mathbf{X}_n$. Then

$$\sum_{i,j=1}^{p} r_{f,1}(i,j)r_{b,1}(i,j) \xrightarrow{P} \operatorname{tr}[\boldsymbol{\Psi}\mathbf{V}\boldsymbol{\Psi}'\mathbf{V}^{-1}],$$

where $\mathbf{V}$ is the covariance matrix of the vector $\mathbf{X}_n$.

*Proof.* Observe that

$$\begin{aligned} \mathbf{C}_1 &= \frac{1}{N}\sum_{n=1}^{N-1}\mathbf{X}_n\mathbf{X}_{n+1}' = \frac{1}{N}\sum_{n=1}^{N-1}\mathbf{X}_n[\boldsymbol{\Psi}\mathbf{X}_n + \boldsymbol{\varepsilon}_{n+1}]' \\ &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{X}_n\mathbf{X}_n'\boldsymbol{\Psi}' + o_P(1) + \frac{1}{N}\sum_{n=1}^{N-1}\mathbf{X}_n\boldsymbol{\varepsilon}_{n+1}' \\ &= \mathbf{V}\boldsymbol{\Psi}' + o_P(1), \end{aligned}$$

by the ergodic theorem. Consequently,

$$\mathbf{C}_0^{-1}\mathbf{C}_1 \xrightarrow{P} \boldsymbol{\Psi}' \qquad \text{and} \qquad \mathbf{C}_1\mathbf{C}_0^{-1} \xrightarrow{P} \mathbf{V}\boldsymbol{\Psi}'\mathbf{V}^{-1},$$

and so $r_{f,1}(i,j) \xrightarrow{P} \psi_{ji}$ and $r_{b,1}(i,j) \xrightarrow{P} [\mathbf{V}\boldsymbol{\Psi}'\mathbf{V}^{-1}]_{ij}$. Therefore,

$$\begin{aligned} \sum_{i,j=1}^{p} r_{f,1}(i,j)r_{b,1}(i,j) &\xrightarrow{P} \sum_{j=1}^{p}\sum_{i=1}^{p}\psi_{ji}[\mathbf{V}\boldsymbol{\Psi}'\mathbf{V}^{-1}]_{ij} \\ &= \sum_{j=1}^{p}[\boldsymbol{\Psi}\mathbf{V}\boldsymbol{\Psi}'\mathbf{V}^{-1}]_{jj} = \operatorname{tr}[\boldsymbol{\Psi}\mathbf{V}\boldsymbol{\Psi}'\mathbf{V}^{-1}]. \end{aligned}$$

*Lemma 2.* If $\mathbf{V}$ is a symmetric positive definite $p \times p$ matrix and $\boldsymbol{\Psi}$ is a nonzero matrix of the same dimension, then $\operatorname{tr}[\boldsymbol{\Psi}\mathbf{V}\boldsymbol{\Psi}'\mathbf{V}^{-1}] > 0$.

*Proof.* To get a feel for why this result is true, suppose that $p = 2$ and $\mathbf{V}$ is diagonal, that is,

$$\mathbf{V} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}.$$

Then

$$\begin{aligned} &\boldsymbol{\Psi}\mathbf{V}\boldsymbol{\Psi}'\mathbf{V}^{-1} \\ &= \begin{bmatrix} \psi_{11}\lambda_1 & \psi_{12}\lambda_2 \\ \psi_{21}\lambda_1 & \psi_{22}\lambda_2 \end{bmatrix}\begin{bmatrix} \psi_{11}\lambda_1^{-1} & \psi_{21}\lambda_2^{-1} \\ \psi_{12}\lambda_1^{-1} & \psi_{22}\lambda_2^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \psi_{11}^2 + \psi_{12}^2\lambda_2\lambda_1^{-1} & \psi_{11}\psi_{21}\lambda_1\lambda_2^{-1} + \psi_{12}\psi_{22} \\ \psi_{21}\psi_{11} + \psi_{22}\psi_{12}\lambda_2\lambda_1^{-1} & \psi_{21}^2\lambda_1\lambda_2^{-1} + \psi_{22}^2 \end{bmatrix}. \end{aligned}$$

Because $\lambda_1$ and $\lambda_2$ are positive, the trace is positive if one of the $\psi_{jk}$'s is positive.

For arbitrary symmetric positive definite $\mathbf{V}$, there is an orthonormal matrix $\mathbf{P}$ such that $\mathbf{V} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$, where $\mathbf{\Lambda}$ is diagonal, and its diagonal consists of positive eigenvalues of $\mathbf{V}$. Therefore,

$$\mathbf{\Psi}\mathbf{V}\mathbf{\Psi}'\mathbf{V}^{-1} = \mathbf{\Psi}\mathbf{P}\mathbf{\Lambda}\mathbf{P}'\mathbf{\Psi}'(\mathbf{P}\mathbf{\Lambda}\mathbf{P}')^{-1} = \mathbf{\Psi}\mathbf{P}\mathbf{\Lambda}\mathbf{P}'\mathbf{\Psi}'\mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}'.$$

Because $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$, setting $\mathbf{A} = \mathbf{\Psi}\mathbf{P}\mathbf{\Lambda}\mathbf{P}'\mathbf{\Psi}'\mathbf{P}\mathbf{\Lambda}^{-1}$ and $\mathbf{B} = \mathbf{P}'$, we obtain

$$\operatorname{tr}[\mathbf{\Psi}\mathbf{V}\mathbf{\Psi}'\mathbf{V}^{-1}] = \operatorname{tr}[\mathbf{P}'\mathbf{\Psi}\mathbf{P}\mathbf{\Lambda}\mathbf{P}'\mathbf{\Psi}'\mathbf{P}\mathbf{\Lambda}^{-1}] = \operatorname{tr}[\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}'\mathbf{\Lambda}^{-1}],$$

where $\mathbf{\Phi} = \mathbf{P}'\mathbf{\Psi}\mathbf{P}$. Because $\mathbf{\Psi} = \mathbf{P}\mathbf{\Phi}\mathbf{P}'$, $\mathbf{\Phi}$ is nonzero. Direct verification shows that the $j$th diagonal entry of $\mathbf{\Phi}\mathbf{V}\mathbf{\Phi}'\mathbf{V}^{-1}$ is $\sum_{k=1}^{p} \phi_{jk}^2 \lambda_k \lambda_j^{-1}$, and so

$$\operatorname{tr}[\mathbf{\Psi}\mathbf{V}\mathbf{\Psi}'\mathbf{V}^{-1}] = \sum_{j=1}^{p}\sum_{k=1}^{p} \phi_{jk}^2 \lambda_k \lambda_j^{-1} > 0. \tag{14}$$

*Corollary 1.* If $\mathbf{V}$ is a symmetric nonnegative definite $p \times p$ matrix and $\mathbf{\Psi}$ is any matrix of the same dimension, then $\operatorname{tr}[\mathbf{\Psi}\mathbf{V}\mathbf{\Psi}'\mathbf{V}] \geq 0$.

*Proof.* Proceeding exactly as in the proof of Lemma 2, we obtain

$$\operatorname{tr}[\mathbf{\Psi}\mathbf{V}\mathbf{\Psi}'\mathbf{V}] = \sum_{j=1}^{p}\sum_{k=1}^{p} \phi_{jk}^2 \lambda_k \lambda_j \geq 0.$$

We now present the remainder of the proof of Theorem 2. Direct verification shows that

$$\sum_{i,j=1}^{p} r_{f,h}^F(i,j) r_{b,h}^F(i,j) = \operatorname{tr}\{[\mathbf{C}_h^F]'[\mathbf{C}_0^F]^{-1}\mathbf{C}_h^F[\mathbf{C}_0^F]^{-1}\}. \tag{15}$$

If $[\mathbf{C}_0^F]^{-1}$ exists, it is positive definite, so, by (14) and Corollary 1,

$$\sum_{i,j=1}^{p} r_{f,h}^F(i,j) r_{b,h}^F(i,j) \geq 0.$$

By Lemmas 1 and 2,

$$\sum_{i,j=1}^{p} r_{f,1}(i,j) r_{b,1}(i,j) \xrightarrow{P} q > 0.$$

It thus suffices to show that

$$\sum_{i,j=1}^{p} [r_{f,1}^F(i,j) r_{b,1}^F(i,j) - r_{f,1}(i,j) r_{b,1}(i,j)] \xrightarrow{P} 0. \tag{16}$$

Relation (16) will follow from $\mathbf{C}_h^F - \mathbf{C}_h \xrightarrow{P} 0$.

In the remainder of the proof, we use the notation introduced in the proof of Theorem 1. We must show that $M_1 \xrightarrow{P} 0$ and $M_2 \xrightarrow{P} 0$. We give only the argument for $M_1$. Observe that

$$M_1 = \left\langle N^{-1} \sum_{n=1}^{N-h} \langle X_{n+h}, v_l\rangle X_n, v_k - v_{kN} \right\rangle.$$

By (3), $\|v_k - v_{kN}\| \xrightarrow{P} 0$. Because

$$E\left\| N^{-1} \sum_{n=1}^{N-h} \langle X_{n+h}, v_l\rangle X_n \right\| \leq E\|\langle X_{n+h}, v_l\rangle X_n\| \leq E\|X_n\|^2,$$

it follows that $M_1 \xrightarrow{P} 0$.

## APPENDIX A: AUXILIARY LEMMAS FOR *H*–VALUED RANDOM ELEMENTS

Consider the empirical lag-$h$ autocovariance operator

$$C_{N,h}(x) = \frac{1}{N} \sum_{n=1}^{N-h} \langle X_n, x\rangle X_{n+h}. \tag{A.1}$$

Recall that the Hilbert–Schmidt norm of a Hilbert–Schmidt operator $S$ is defined by

$$\|S\|_{\mathcal{S}}^2 = \sum_{j=1}^{\infty} \|S(e_j)\|^2,$$

where $\{e_1, e_2, \ldots\}$ is any orthonormal basis.

*Lemma A.1.* Suppose that the $X_i$'s are iid random elements in a separable Hilbert space with $E\|X_0\|^2 < \infty$; then

$$E\|C_{N,h}\|_{\mathcal{S}}^2 = \frac{N-h}{N^2}(E\|X_0\|^2)^2.$$

*Proof.* Observe that

$$\begin{aligned}
\|C_{N,h}\|_{\mathcal{S}}^2 &= \sum_{j=1}^{\infty} \|C_{N,h}(e_j)\|^2 \\
&= \sum_{j=1}^{\infty} \left\langle \frac{1}{N}\sum_{n=1}^{N-h} \langle X_n, e_j\rangle X_{n+h}, \frac{1}{N}\sum_{m=1}^{N-h} \langle X_m, e_j\rangle X_{m+h} \right\rangle \\
&= \sum_{j=1}^{\infty} \frac{1}{N^2} \sum_{m,n=1}^{N-h} \langle X_m, e_j\rangle\langle X_n, e_j\rangle\langle X_{m+h}, X_{n+h}\rangle.
\end{aligned}$$

It follows from the independence of the $X_n$ that

$$\begin{aligned}
E\|C_{N,h}\|_{\mathcal{S}}^2 &= \frac{1}{N^2} \sum_{n=1}^{N-h}\sum_{j=1}^{\infty} E[\langle X_n, e_j\rangle]^2 E[\langle X_{n+h}, X_{n+h}\rangle]^2 \\
&= E\|X_0\|^2 \frac{1}{N^2} \sum_{n=1}^{N-h} E\left[\sum_{j=1}^{\infty} [\langle X_n, e_j\rangle]^2\right] \\
&= [E\|X_0\|^2]^2 \frac{N-h}{N^2}.
\end{aligned}$$

*Lemma A.2.* Suppose that $X_n$, $Z_N$, and $Y_N$ are random elements in a separable Hilbert space. We assume that

$$E\|Y_N\|^2 = O(N^{-1}), \qquad E\|Z_N\|^2 = O(N^{-1}) \tag{A.2}$$

and

$$X_n \sim iid, \qquad E\|X_n\|^2 < \infty. \tag{A.3}$$

Then

$$N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, Y_N\rangle\langle X_{n+h}, Z_N\rangle \xrightarrow{P} 0.$$

*Proof.* Observe that

$$N^{-1/2}\sum_{n=1}^{N-h}\langle X_n, Y_N\rangle\langle X_{n+h}, Z_N\rangle = \langle C_{N,h}(Y_N), N^{1/2}Z_N\rangle,$$

with the operator $C_{N,h}$ defined in (A.1). Because $P(N^{1/2}\|Z_N\| > C) \le C^{-2}NE\|Z_N\|^2$, $N^{1/2}\|Z_N\| = O_P(1)$. Thus it remains to verify that $C_{N,h}(Y_N) \xrightarrow{P} 0$. Because the Hilbert–Schmidt norm is not less than the uniform operator norm $\|\cdot\|_{\mathcal{L}}$ (see Bosq 2000, p. 35), we obtain, from Lemma A.1,

$$\begin{aligned} E\|C_{N,h}(Y_N)\| &\le E[\|C_{N,h}\|_{\mathcal{L}}\|Y_N\|] \le E[\|C_{N,h}\|_{\mathcal{S}}\|Y_N\|] \\ &\le (E\|C_{N,h}\|_{\mathcal{S}}^2)^{1/2}(E\|Y_N\|^2)^{1/2} \\ &= O\big(N^{-1/2}\big)O\big(N^{-1/2}\big) = O(N^{-1}). \end{aligned}$$

## APPENDIX B: LIMIT THEORY FOR SAMPLE AUTOCOVARIANCE MATRICES

Detailed proofs of Theorems B.1, B.2, and B.3 and Remark B.1, and all displayed formulas leading to them, are omitted to conserve space but are available on request. Consider random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_N$, where $\mathbf{X}_t = [X_{1t}, X_{2t}, \ldots, X_{pt}]'$. We assume that the $\mathbf{X}_t$, $t = 1, 2, \ldots$, are iid mean 0 with finite variance and write

$$v(i,j) = E[X_{it}X_{jt}] \qquad \text{and} \qquad \mathbf{V} = [v(i,j)]_{i,j=1,\ldots,p}.$$

We let $\mathbf{C}_h$ denote the sample autocovariance matrix with entries

$$c_h(k,l) = \frac{1}{N}\sum_{t=1}^{N-h} X_{kt}X_{l,t+h}, \qquad h \ge 0.$$

We first establish the joint asymptotic distribution of $\mathbf{C}_h$, $h = 1, 2, \ldots, H$. We use the asymptotic normality of $h$-independent sequences (see thm. 6.4.2 of Brockwell and Davis 1991), and the Wald–Cramer device.

Let $\mathbf{Z}_0$ and $\mathbf{Z}_h$ be matrices with jointly Gaussian entries $Z_0(k,l)$, $Z_h(k,l)$, $k,l = 1, \ldots, p$, with mean 0 and covariances

$$E[Z_0(k,l)Z_0(i,j)] = \eta(k,l,i,j) - v(i,j)v(k,l) \tag{B.1}$$

and

$$E[Z_h(k,l)Z_h(i,j)] = v(k,i)v(l,j) \qquad (h \ge 1), \tag{B.2}$$

where $\eta(k,l,i,j) = E[X_{kt}X_{lt}X_{it}X_{jt}]$.

*Theorem B.1.* If the $\mathbf{X}_t$'s are iid with finite fourth moment, then

$$N^{1/2}[\mathbf{C}_0 - V, \mathbf{C}_1, \ldots, \mathbf{C}_H] \xrightarrow{d} [\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_H],$$

where the $\mathbf{Z}_h$, $h = 0, 1, \ldots, H$, are independent mean-0 Gaussian matrices with covariances (B.1) and (B.2).

A critical ingredient of the derivation of the asymptotic distribution of the test statistic $Q_N$ is the understanding of the asymptotic distribution of $\mathbf{C}_0^{-1}$. Let $u(k,l)$ be the $(k,l)$ entry of $\mathbf{V}^{-1}$. Calculations involving derivatives of products of matrices and the delta method lead to the following result:

$$N^{1/2}(\mathbf{C}_0^{-1} - \mathbf{V}^{-1}) \xrightarrow{d} \mathbf{Y}_0, \tag{B.3}$$

where $\mathbf{Y}_0$ is a mean-0 Gaussian matrix with $(i,j)$ entry

$$Y_0(i,j) = -\sum_{k,l=1}^{p} u(i,k)u(l,j)Z_0(k,l). \tag{B.4}$$

*Remark B.1.* Observe that

$$\begin{aligned} &E[Y_0(i,j)Y_0(\alpha,\beta)] \\ &\quad = \sum_{k,l=1}^{p}\sum_{\kappa,\lambda=1}^{p} u(i,k)u(l,j)u(\alpha,\kappa)u(\lambda,\beta)E[Z_0(k,l)Z_0(\kappa,\lambda)]. \end{aligned}$$

The covariances $E[Z_0(k,l)Z_0(\kappa,\lambda)]$ are given in (B.1), and it is seen that they do not imply, in general, the covariances in formula (4.4) of Chitturi (1976), which is true only if the process $\mathbf{X}_t$ is Gaussian.

We now find the limit of $N^{1/2}\mathbf{C}_0^{-1}\mathbf{C}_h$, $h \ge 1$. Further calculations using the delta method applied to the matrix $[\mathbf{C}_0 - \mathbf{V}, \mathbf{C}_1, \ldots, \mathbf{C}_H]'$ lead to the following theorem.

*Theorem B.2.* If the $\mathbf{X}_t$'s are iid with finite fourth moment, then

$$N^{1/2}\mathbf{C}_0^{-1}[\mathbf{C}_1, \ldots, \mathbf{C}_H] \xrightarrow{d} \mathbf{V}^{-1}[\mathbf{Z}_1, \ldots, \mathbf{Z}_H], \tag{B.5}$$

where the $\mathbf{Z}_h$, $h = 0, 1, \ldots, H$, are independent mean-0 Gaussian matrices with covariances (B.2).

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

$$Q_N = N\sum_{h=1}^{H}\sum_{i,j=1}^{p} r_{f,h}(i,j)r_{b,h}(i,j). \tag{B.6}$$

*Theorem B.3.* If the $\mathbf{X}_t$'s are iid with finite fourth moment, then $Q_N \xrightarrow{d} \chi^2_{p^2H}$.

*Proof.* Similarly to (B.5), it can be verified that

$$N^{1/2}[\mathbf{C}_1, \ldots, \mathbf{C}_H]\mathbf{C}_0^{-1} \xrightarrow{d} [\mathbf{Z}_1, \ldots, \mathbf{Z}_H]\mathbf{V}^{-1}, \tag{B.7}$$

and that convergence (B.5) and (B.7) are joint. Because the matrices $[\mathbf{C}_0^{-1}\mathbf{C}_h, \mathbf{C}_h\mathbf{C}_0^{-1}]$ are asymptotically independent, it suffices to verify that

$$N\sum_{i,j=1}^{p} r_{f,h}(i,j)r_{b,h}(i,j) \xrightarrow{d} \chi^2_{p^2}. \tag{B.8}$$

To lighten the notation, in the remainder of the proof we suppress the index $h$ (the limit distributions do not depend on $h$).

Let $\rho_f(i,j)$ and $\rho_b(i,j)$ denote the entries of matrices $\mathbf{V}^{-1}\mathbf{Z}$ and $\mathbf{Z}\mathbf{V}^{-1}$. By (B.5) and (B.7), it suffices to show that

$$\sum_{i,j=1}^{p} \rho_f(i,j)\rho_b(i,j) \overset{d}{=} \chi^2_{p^2}. \tag{B.9}$$

Let $\tilde{\mathbf{Z}}$ denote the column vector of length $p^2$ obtained by expanding the matrix $\mathbf{Z}$ row by row. Then the covariance matrix of $\tilde{\mathbf{Z}}$ is the $p^2 \times p^2$ matrix $\mathbf{V} \otimes \mathbf{V}$. By formula (23) of Anderson (1984, p. 600), its inverse is $(\mathbf{V} \otimes \mathbf{V})^{-1} = \mathbf{V}^{-1} \otimes \mathbf{V}^{-1} = \mathbf{U} \otimes \mathbf{U}$. It thus follows from theorem 3.3.3 of Anderson (1984) that

$$\tilde{\mathbf{Z}}'(\mathbf{U} \otimes \mathbf{U})\tilde{\mathbf{Z}} \overset{d}{=} \chi^2_{p^2}. \tag{B.10}$$

It remains to show that the left side of (B.9) is equal to the left side of (B.10). The entry $Z(i,k)$ of the vector $\tilde{\mathbf{Z}}'$ multiplies the row $u(i,\cdot)u(k,\cdot)$ of $\mathbf{U} \otimes \mathbf{U}$; the entry $Z(j,l)$ of $\tilde{\mathbf{Z}}$ multiplies the column $u(\cdot,j)u(\cdot,l)$. Consequently,

$$\begin{aligned} \tilde{\mathbf{Z}}'(\mathbf{U} \otimes \mathbf{U})\tilde{\mathbf{Z}} &= \sum_{i,j,k,l=1}^{p} u(i,j)u(k,l)Z(i,k)Z(j,l) \\ &= \sum_{i,l=1}^{p}\sum_{j=1}^{p} u(i,j)Z(j,l)\sum_{k=1}^{p} Z(i,k)u(k,l) \end{aligned}$$

$$= \sum_{i,l=1}^{p} \rho_f(i,l)\rho_b(i,l),$$

completing the proof.

*[Received March 2007. Revised July 2007.]*

## REFERENCES


Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

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

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer-Verlag.

Chitturi, R. V. (1976), "Distribution of Multivariate White Noise Autocorrelation," *Journal of the American Statistical Association*, 71, 223–226.

Fernández de Castro, B., Guillas, S., and Gonzáles Manteiga, W. (2005), "Functional Samples and Bootstrap for Predicting Sulfur Dioxide Levels," *Technometrics*, 47, 212–222.

Hosking, J. R. M. (1980), "The Multivariate Portmanteau Statistic," *Journal of the American Statistical Association*, 75, 602–608.

Jach, A., Kokoszka, P., Sojka, J., and Zhu, L. (2006), "Wavelet–Based Index of Magnetic Storm Activity," *Journal of Geophysical Research*, 111, A09215.

Kargin, V., and Onatski, A. (2007), "Curve Forecasting by Functional Autoregression," technical report, New York University, Mathematics Dept./ Columbia University, Economics Dept.

Laukaitis, A., and Račkauskas, A. (2002), "Functional Data Analysis of Payment Systems," *Nonlinear Analysis: Modeling and Control*, 7, 53–68.

Li, W. K. (1981), "Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models," *Journal of the Royal Statistical Society*, Ser. B, 43, 231–239.

Ljung, G., and Box, G. (1978), "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 66, 67–72.

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

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

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

——— (2005), *Functional Data Analysis*, New York: Springer-Verlag.

Seber, G. A. F. (1984), *Multivariate Observations*, New York: Wiley.

Xu, W.-Y., and Kamide, Y. (2004), "Decomposition of Daily Geomagnetic Variations by Using Method of Natural Orthogonal Component," *Journal of Geophysical Research*, 109, A05218.