Journal of Multivariate Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Testing the stability of the functional autoregressive process

Lajos Horváth [a], Marie Hušková [b], Piotr Kokoszka [c,*]

[a] Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City, UT 84112-0090, USA

[b] Department of Statistics, Charles University, Sokolovská 83, CZ-18600 Praha, Czech Republic

[c] Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, USA

ARTICLE INFO

Article history:
Available online xxxx

AMS subject classification:
62M10

Keywords:
Change-point
Functional autoregressive process

ABSTRACT

The functional autoregressive process has become a useful tool in the analysis of functional time series data. It is defined by the equation $X_{n+1} = \mathbf{\Psi} X_n + \varepsilon_{n+1}$, in which the observations X_n and errors ε_n are curves, and $\mathbf{\Psi}$ is an operator. To ensure meaningful inference and prediction based on this model, it is important to verify that the operator $\mathbf{\Psi}$ does not change with time. We propose a method for testing the constancy of $\mathbf{\Psi}$ against a change-point alternative which uses the functional principal component analysis. The test statistic is constructed to have a well-known asymptotic distribution, but the asymptotic justification of the procedure is very delicate. We develop a new truncation approach which together with Mensov's inequality can be used in other problems of functional time series analysis. The estimation of the principal components introduces asymptotically non-negligible terms, which however cancel because of the special form of our test statistic (CUSUM type). The test is implemented using the R package `fda`, and its finite sample performance is examined by application to credit card transaction data.

1. Introduction

Functional data analysis (FDA) has enjoyed increased popularity over the last decade due to its applicability to problems which are difficult to cast into a framework of scalar or vector observations. Even if such standard approaches are available, the functional approach often leads to a more natural and parsimonious description of the data, and to more accurate inference and prediction. The monograph of Ramsay and Silverman [1] has become a standard reference to the ideas and tools of FDA. To name a few recent applications of FDA which illustrate its advantages alluded to above, we cite [2–4], and the recent monograph of Ferraty and Vieu [5].

Functional time series (FTS) arise when a long record $\{X(t),\ t \in [0, T]\}$, in which t is, at least conceptually, a continuous (real) index, can be naturally split into segments of equal, say unit, length. We then set $X_n(t) = X(n-1+t),\ t \in [0, 1],\ n = 1, 2, \ldots N = T$. The data are then the curves $X_n(\cdot),\ n = 1, 2, \ldots N$. Seven consecutive FTS observations are shown in Fig. 1, the whole time series is described and analyzed in Section 4.

The simplest model for a FTS is the ARH(1) model of Bosq [6], which extends to the functional setting the usual AR(1) model. Despite its conceptual simplicity, it is a very flexible modeling and predictive tool because the autoregressive operator acts on a Hilbert space whose elements can exhibit any degree of nonlinearity. Thus, even though ARH(1) is a linear model in a function space, it is highly nonlinear for individual scalar records. Various nonparametric estimation and prediction methods for the ARH(1) model have been put forward, and it has found numerous applications, see [7–12], among others.

In contrast to functional data derived from designed experiments, for FTS it is important to verify if a single model can be used for the whole record. Conditions may change with time, leading to a break in the stochastic structure of the data.

* Corresponding author.
E-mail address: Piotr.Kokoszka@usu.edu (P. Kokoszka).

0047-259X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2008.12.008

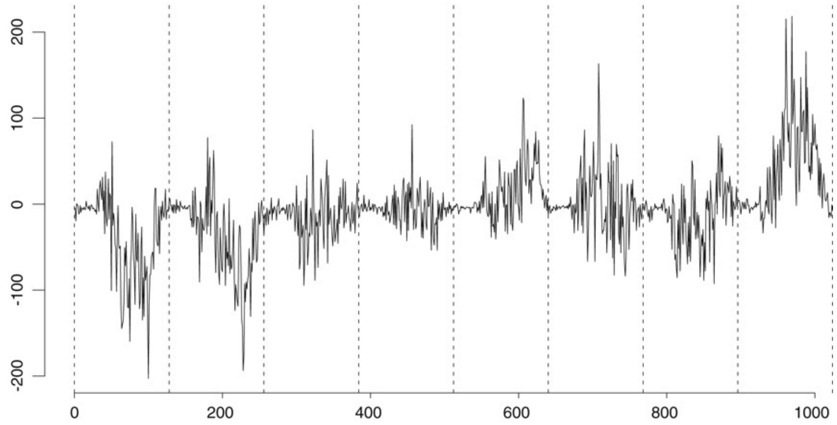

Fig. 1. Seven observations (days) of a functional time series derived from credit card transaction data. The vertical dotted lines separate days.

Failure to take such change points into account leads to erroneous inference. In this paper, we propose a test of the stability of the ARH(1) model against a change point alternative. This problem can be stated formally as follows. We observe the random functions $\{X_n(t),\ t \in [0,1],\ n = 1, 2, \ldots N\}$ and assume that they follow the model

$$X_{n+1} = \Psi_n X_n + \varepsilon_{n+1}, \quad n = 1, 2, \ldots, N, \tag{1.1}$$

with independent identically distributed (iid) mean zero innovations $\varepsilon_n \in L^2([0,1])$. Precise conditions on the operators Ψ_n are formulated in Section 2. We want to test

$$H_0 : \Psi_1 = \Psi_2 = \cdots = \Psi_N$$

against the alternative

$$H_A : \text{there is } 1 \le k^* < N : \Psi_1 = \cdots = \Psi_{k^*} \ne \Psi_{k^*+1} = \cdots = \Psi_N.$$

Under H_0, the common operator is denoted by Ψ.

The test statistic is based on the differences of the sample autocovariances of projections of the functional observations on estimated principal components (PC's). The limit distribution can be derived by replacing the estimated PC's by their population counterparts and using a functional central limit theorem for ergodic sequences. But in the functional setting, this replacement introduces asymptotically nonnegligible terms, see Section 7, which cancel because of the special form of the test statistic (see also [13]). The estimated PC's are determined only up to a sign, and our statistic is invariant to these random signs, see Section 4. Finally, to show that the remaining terms due to the estimation of the PC's are asymptotically negligible, we develop a new technique which involves the truncation at lag $O(\log N)$ of the moving average representation of the ARH(1) process (Lemma 7.2), a blocking technique that utilizes this truncation (Lemma 7.3) and Mensov's inequality (Lemma 7.7). We hope that these tools will prove useful in other inference problems related to the functional ARH(1) model.

A somewhat related problem is studied in [14] who considered two populations, admitting the PCA's:

$$X_{i,p}(t) = \mu_p(t) + \sum_{1 \le \ell < \infty} \eta_{i,p,\ell}\phi_{p,\ell}(t), \quad p = 1, 2.$$

Benko et al. [14] developed a bootstrap test for checking if the elements of the two decompositions are the same. In our setting, we do not have a specific partition into two sets. The change can occur at any point, and we want to test if it occurs or not. We do not test for the change in the structure of the PC's of iid observations, but in the dependence structure of a FTS.

Testing for a change point in the mean function of iid functional observation is considered in [15]. An informal way of testing the stability of the ARH(1) model by a visual examination of the scores is performed in [12]. Laukaitis [16] studies the residuals $\hat{\varepsilon}_{n+1} = X_{n+1} - \hat{\Psi} X_n$ in a Hilbert Space, and argues that they could be used for change-point detection.

The paper is organized as follows. Section 2 introduces the relevant notation and assumptions. The testing procedure is described and heuristically justified in Section 3. Its application and finite sample performance is examined in Section 4. Asymptotic justification is presented in Section 5, with the proofs developed in Sections 6 and 7.

2. Preliminaries

To focus attention, we assume that all random functions are defined on the interval [0, 1], but our theory remains valid if it is replaced by a compact subset of a Euclidean space. To lighten the notation, we do not indicate the limits of integration, i.e. we use $\int f(t)\mathrm{d}t$ to denote $\int_0^1 f(t)\mathrm{d}t$.

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

The following assumption formalizes the structure of the observations under the null hypothesis.

Assumption 2.1. The functional observations $X_n \in L^2([0, 1])$ satisfy

$$X_{n+1} = \Psi X_n + \varepsilon_{n+1}, \quad n = 0, 1, \ldots, N-1, \tag{2.1}$$

where Ψ is an integral operator with the kernel $\psi(t, s)$ satisfying

$$\int\int \psi^2(s, t)\mathrm{d}s\mathrm{d}t < 1, \tag{2.2}$$

and the iid mean zero innovations $\varepsilon_n \in L^2([0, 1])$ satisfy

$$E\|\varepsilon_n\|^4 = E\left[\int \varepsilon_n^2(t)\mathrm{d}t\right]^2 < \infty. \tag{2.3}$$

Eq. (2.1) can be written more explicitly as

$$X_{n+1}(t) = \int \psi(t, s)X_n(s)\mathrm{d}s + \varepsilon_{n+1}(t), \quad t, s \in [0, 1]. \tag{2.4}$$

Assumption 2.1 ensures that (2.4) has a unique strictly stationary solution $\{X_n(t),\ t \in [0, 1]\}$ with finite fourth moment in $L^2([0, 1])$ such that ε_{n+1} is independent of $X_n, X_{n-1}, \ldots$, see Chapter 3 of [6]. In our context it is not enough to assume that the noise sequence is a weak white noise with finite second moment. We need the assumptions of independence and finite fourth moment to be able to use the estimation results developed in Chapter 4 of [6] and other CLT type results.

If Assumption 2.1 holds, we can define the covariance operator

$$C(x) = E[\langle X_n, x\rangle X_n], \quad x \in L^2([0, 1])$$

whose eigenfunctions v_j and eigenvalues λ_j are defined by

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

The eigenfunctions v_j form an orthonormal basis of $L^2([0, 1])$.

The empirical covariance operator is defined by

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

and its eigenelements by

$$\hat{C}(\hat{v}_j) = \hat{\lambda}_j \hat{v}_j, \quad j = 1, 2, \ldots N.$$

Since the operators C and $\hat{C}$ are symmetric and nonnegative definite, the eigenvalues λ_j and $\hat{\lambda}_j$ are nonnegative.

In order to develop an asymptotic justification for our procedure, we need a bound on the distance between the estimated eigenelements $\hat{\lambda}_j$, $\hat{v}_j$ and their population counterparts λ_j, v_j. If the eigenspace corresponding to the eigenvalue λ_k is one dimensional, then formulas (4.39) and (4.49) of [6] imply, with $\hat{c}_k = \langle \hat{v}_k, v_k\rangle$,

$$\limsup_{N\to\infty} NE\|\lambda_k - \hat{\lambda}_k\|^2 < \infty, \qquad \limsup_{N\to\infty} NE\|v_k - \hat{c}_k\hat{v}_k\|^2 < \infty. \tag{2.5}$$

(Note that n before the expected value is missing in formula (4.49) of [6], cf. formulas (4.17) and (4.44) of that monograph.) For this reason, we impose the following additional assumption.

Assumption 2.2. The eigenvalues of the covariance operator C satisfy

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

3. Testing procedure

In this section, we describe the idea of the test and explain its practical application. The requisite asymptotic theory is presented in Section 5.

The idea is to check if the action of Ψ on the span of the p most important principal components of the observations $X_1, X_2, \ldots, X_N$ changes at some unknown time point i. If there is no change in the autoregressive operator Ψ, the functions Ψv_j, $j = 1, 2, \ldots, p$, remain constant. Since $\Psi v_j = \sum_\ell \langle \Psi v_j, v_\ell\rangle v_\ell$, this is the case, to a good approximation, if the coefficients $\langle \Psi v_j, v_\ell\rangle$, $\ell \leq p$ remain constant. Direct verification shows that under H_0, $\langle \Psi v_j, v_\ell\rangle = \lambda_j^{-1}\langle Rv_j, v_\ell\rangle$ where

$$Rx = E[\langle X_n, x\rangle X_{n+1}]$$

is the lag-1 autocovariance operator. Thus, the constancy of Ψ is approximately equivalent to the constancy of the products $\langle Rv_j, v_\ell\rangle$, $j, \ell = 1, 2, \ldots, p$.

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

The restriction to the action of Ψ on the span of $v_j,\ j,=1,2,\ldots,p$, means that the test will not detect changes on the orthogonal complement of this space. Typically p is chosen so that the empirical counterparts $\hat{v}_j,\ j,=1,2,\ldots,p$, explain a large percentage of the variability of the data and their linear combinations approximate the data very closely, see Chapter 8 of [1]. We therefore view a change in the action of Ψ on $v_j, j>p$, as not relevant. This restriction quantifies the intuition that very small changes cannot be detected. Another point to note is that since $\langle Rv_j, v_\ell\rangle = \lambda_j \langle \Psi v_j, v_\ell\rangle$, a change in Ψ may be obscured by a change in the eigenfunctions λ_j, thus potentially reducing power. Nevertheless, the test introduced below is effective in practical settings, and its large sample properties are tractable.

To devise a test against the alternative of a change-point, we must first estimate these products from observations $X_1, X_2, \ldots, X_k$, then from observations $X_{k+1}, X_{k+2}, \ldots, X_N$, and compare the resulting estimates. To achieve it, we define p-dimensional projections

$$\mathbf{X}_i = [X_{i1}, \ldots, X_{ip}]^{\mathrm{T}}, \qquad \hat{\mathbf{X}}_i = [\hat{X}_{i1}, \ldots, \hat{X}_{ip}]^{\mathrm{T}}$$

by

$$X_{ij} = \langle X_i, v_j\rangle = \int X_i(t)v_j(t)\mathrm{d}t, \qquad \hat{X}_{ij} = \langle X_i, \hat{v}_j\rangle = \int X_i(t)\hat{v}_j(t)\mathrm{d}t$$

and $p \times p$ matrices

$$\mathbf{R}_k = \frac{1}{k}\sum_{2\le i\le k} \mathbf{X}_{i-1}\mathbf{X}_i^{\mathrm{T}}, \qquad \mathbf{R}^*_{N-k} = \frac{1}{N-k}\sum_{k<i\le N} \mathbf{X}_{i-1}\mathbf{X}_i^{\mathrm{T}};$$

$$\hat{\mathbf{R}}_k = \frac{1}{k}\sum_{2\le i\le k} \hat{\mathbf{X}}_{i-1}\hat{\mathbf{X}}_i^{\mathrm{T}}, \qquad \hat{\mathbf{R}}^*_{N-k} = \frac{1}{N-k}\sum_{k<i\le N} \hat{\mathbf{X}}_{i-1}\hat{\mathbf{X}}_i^{\mathrm{T}}.$$

Observe that by the ergodic theorem, as $k \to \infty$,

$$\mathbf{R}_k(j,\ell) = \frac{1}{k}\sum_{2\le i\le k} \langle X_{i-1}, v_j\rangle \langle X_i, v_\ell\rangle \xrightarrow{a.s.} E[\langle X_{n-1}, v_j\rangle \langle X_n, v_\ell\rangle] = \langle Rv_j, v_\ell\rangle .$$

Thus the matrices $\mathbf{R}_k$ and $\mathbf{R}^*_{N-k}$ approximate the matrix $[\langle Rv_j, v_\ell\rangle,\ j,\ell = 1,2,\ldots,p]$ based, correspondingly, on the observations before and after time k, and so it is appealing to base the test on their difference. The matrices $\mathbf{R}_k$ and $\mathbf{R}^*_{N-k}$ cannot however be computed from the data because the population principal components v_j are unknown. Thus, we would like to replace them by their empirical counterparts $\hat{\mathbf{R}}_k$ and $\hat{\mathbf{R}}^*_{N-k}$. This is however a delicate point because it cannot be guaranteed that $\hat{v}_j$ is close to v_j. Relation (2.5) means that $\hat{c}_j\hat{v}_j$ is close to v_j. Consequently, the (j,ℓ) entry of $\hat{\mathbf{R}}_k$ must be multiplied by $\hat{c}_j\hat{c}_\ell$ in order to approximate the (j,ℓ) entry of $\mathbf{R}_k$. The random signs $\hat{c}_j$ and $\hat{c}_\ell$ are unknown, so a test statistic must be constructed in such a way that they do not appear in it. This is not a mere technical point. Working with the R package `fda`, we have noticed that changing just a few observations can flip the curves $\hat{v}_j$. Moreover, for data with complex multiscale features, arising, for example, in transient geophysical processes, see e.g. [17,18], the interpretation of the empirical principal components is difficult, and their sign is unstable from sample to sample.

We now describe how to construct a number of test statistics which do not depend on the signs $\hat{c}_j$ and $\hat{c}_\ell$, and postpone the rigorous verification to Section 5.

Denote

$$Y_i(j,\ell) = \langle X_{i-1}, v_j\rangle \langle X_i, v_\ell\rangle, \qquad \hat{Y}_i(j,\ell) = \langle X_{i-1}, \hat{v}_j\rangle \langle X_i, \hat{v}_\ell\rangle \tag{3.1}$$

and consider the column vectors of length p^2:

$$\mathbf{Y}_i = [Y_i(1,1),\ldots,Y_i(1,p),Y_i(2,1),\ldots,Y_i(2,p),\ldots,Y_i(p,1),\ldots,Y_i(p,p)]^{\mathrm{T}};$$
$$\hat{\mathbf{Y}}_i = [\hat{Y}_i(1,1),\ldots,\hat{Y}_i(1,p),\hat{Y}_i(2,1),\ldots,\hat{Y}_i(2,p),\ldots,\hat{Y}_i(p,1),\ldots,\hat{Y}_i(p,p)]^{\mathrm{T}}.$$

Define further

$$\mathbf{Z}_k = \sum_{2\le i\le k} \mathbf{Y}_i, \qquad \mathbf{Z}^*_{N-k} = \sum_{k<i\le N} \mathbf{Y}_i;$$
$$\hat{\mathbf{Z}}_k = \sum_{2\le i\le k} \hat{\mathbf{Y}}_i, \qquad \hat{\mathbf{Z}}^*_{N-k} = \sum_{k<i\le N} \hat{\mathbf{Y}}_i.$$

Since the X_i follow a functional AR(1) model, the vectors $\mathbf{Y}_i$ form a weakly dependent stationary sequence, and so, as $k \to \infty$,

$$\sqrt{k}\left[\frac{1}{k}\sum_{2\le i\le k} \mathbf{Y}_i - E\,\mathbf{Y}_k\right] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{D}), \tag{3.2}$$

where $\boldsymbol{D}$ is the $p^2 \times p^2$ long run covariance matrix defined by

$$\boldsymbol{D} = E[(\mathbf{Y}_0 - E\mathbf{Y}_0)(\mathbf{Y}_0 - E\mathbf{Y}_0)^{\mathrm{T}}] + 2 \sum_{1 \le h < \infty} E[(\mathbf{Y}_0 - E\mathbf{Y}_0)(\mathbf{Y}_h - E\mathbf{Y}_h)^{\mathrm{T}}]. \tag{3.3}$$

Relation (3.2), and the corresponding relation for the sum over $k < i \le N$, can be rewritten as

$$\mathbf{Z}_k - kE\mathbf{Y}_N \approx N(0, k\boldsymbol{D}), \qquad \mathbf{Z}^*_{N-k} - (N-k)E\mathbf{Y}_N \approx N(0, (N-k)\boldsymbol{D}).$$

Denoting by $\{\mathbf{W}_D(t),\ t \ge 0\}$ a p^2-dimensional Brownian motion with covariance matrix $\boldsymbol{D}$, we have, in fact,

$$\mathbf{Z}_k - kE\mathbf{Y}_N \approx \mathbf{W}_D(k), \qquad \mathbf{Z}^*_{N-k} - (N-k)E\mathbf{Y}_N \approx \mathbf{W}_D(N) - \mathbf{W}_D(k). \tag{3.4}$$

By (3.4), under H_0 we have,

$$\begin{aligned}\frac{1}{k}\mathbf{Z}_k - \frac{1}{N-k}\mathbf{Z}^*_{N-k} &\approx \frac{1}{k}\mathbf{W}_D(k) - \frac{1}{N-k}\left(\mathbf{W}_D(N) - \mathbf{W}_D(k)\right)\\ &= \frac{1}{k(N-k)}\left[N\mathbf{W}_D(k) - k\mathbf{W}_D(k) - k\mathbf{W}_D(N) + k\mathbf{W}_D(k)\right]\\ &= \frac{N}{k(N-k)}\left[\mathbf{W}_D(k) - \frac{k}{N}\mathbf{W}_D(N)\right].\end{aligned}$$

Denote

$$\mathbf{U}_N(k) = \frac{k(N-k)}{N}\left(\frac{1}{k}\mathbf{Z}_k - \frac{1}{N-k}\mathbf{Z}^*_{N-k}\right). \tag{3.5}$$

The above calculation shows that

$$\mathbf{U}_N(k) \approx \mathbf{W}_D(k) - \frac{k}{N}\mathbf{W}_D(N).$$

Comparing covariances, we see that

$$\frac{1}{N}\left[\mathbf{W}_D(k) - \frac{k}{N}\mathbf{W}_D(N)\right]^{\mathrm{T}} \boldsymbol{D}^{-1}\left[\mathbf{W}_D(k) - \frac{k}{N}\mathbf{W}_D(N)\right], \quad 1 \le k \le N,$$

has the same distribution as

$$\sum_{1 \le m \le p^2} B_m^2(k/N), \quad 1 \le k \le N, \tag{3.6}$$

where the $B_m(\cdot)$ are independent standard Brownian bridges on $[0, 1]$. Consequently, any functional of

$$G_N(k) = \frac{1}{N}\mathbf{U}_N(k)^{\mathrm{T}}\boldsymbol{D}^{-1}\mathbf{U}_N(k), \quad 1 \le k \le N, \tag{3.7}$$

can be approximated by the corresponding functional of (3.6).

Asymptotic theory for functionals of the process $\sum_{1 \le m \le d} B_m^2(u),\ u \in [0, 1]$, including weighted sums and maximally selected statistics, is well-known, see e.g. [19,20], and goes back to [21]. A Cramér–von–Mises type functional $K_d := \int_0^1 \sum_{1 \le m \le d} B_m^2(u)\mathrm{d}u$ leads to tests with good finite sample properties, and so we focus on it in the following, but clearly other functionals can be used as well, see e.g. [22] for more examples.

To implement the test, we need to estimate the matrix $\boldsymbol{D}$ in (3.3). The estimation of the long run covariance matrix is one of the most extensively studied topics in time series analysis and econometrics, see e.g. [23–25] for recent approaches and references. Any reasonable method can be used, but for concreteness, we focus on the popular and simple Bartlett estimator, and explain how to adapt it to the change point problem.

Denote by

$$\widehat{\boldsymbol{\gamma}}_h(k) = \frac{1}{k}\sum_{1 \le i \le k-h}\left(\hat{\mathbf{Y}}_i - \frac{1}{k}\sum_{1 \le i \le k}\hat{\mathbf{Y}}_i\right)\left(\hat{\mathbf{Y}}_{i+h} - \frac{1}{k}\sum_{1 \le i \le k}\hat{\mathbf{Y}}_i\right)^{\mathrm{T}}$$

and

$$\widehat{\boldsymbol{\gamma}}^*_h(N-k) = \frac{1}{N-k}\sum_{k < i \le N-h}\left(\hat{\mathbf{Y}}_i - \frac{1}{N-k}\sum_{k < i \le N-h}\hat{\mathbf{Y}}_i\right)\left(\hat{\mathbf{Y}}_{i+h} - \frac{1}{N-k}\sum_{k < i \le N-h}\hat{\mathbf{Y}}_i\right)^{\mathrm{T}}$$

the lag h $p^2 \times p^2$ autocovariance matrices computed, respectively, from the first k and the last $N-k$ observations. The corresponding Bartlett estimators of $\boldsymbol{D}$ are then

$$\widehat{\boldsymbol{D}}_k = \widehat{\boldsymbol{\gamma}}_0(k) + 2\sum_{1\le h\le q}\left(1-\frac{h}{q+1}\right)\widehat{\boldsymbol{\gamma}}_h(k) \tag{3.8}$$

and

$$\widehat{\boldsymbol{D}}^*_{N-k} = \widehat{\boldsymbol{\gamma}}^*_0(N-k) + 2\sum_{1\le h\le q}\left(1-\frac{h}{q+1}\right)\widehat{\boldsymbol{\gamma}}^*_h(N-k). \tag{3.9}$$

The sequence $G_N(k)$ (3.7) is approximated by the sequence

$$\hat{G}_N(k) = \frac{1}{N}\hat{\mathbf{U}}_N(k)^{\mathrm{T}}\left[\frac{k}{N}\widehat{\boldsymbol{D}}_k + \left(1-\frac{k}{N}\right)\widehat{\boldsymbol{D}}^*_{N-k}\right]^{-1}\hat{\mathbf{U}}_N(k), \tag{3.10}$$

where

$$\hat{\mathbf{U}}_N(k) = \frac{k(N-k)}{N}\left(\frac{1}{k}\hat{\mathbf{Z}}_k - \frac{1}{N-k}\hat{\mathbf{Z}}^*_{N-k}\right). \tag{3.11}$$

Using the weighted sum of the estimators $\widehat{\boldsymbol{D}}_k$ and $\widehat{\boldsymbol{D}}^*_{N-k}$ in (3.10) has been shown in different settings to lead to better power than using just $\widehat{\boldsymbol{D}}_N$, see [26,27].

Defining the critical value $c(\alpha, d)$ by $P(K_d > c(\alpha, d)) = \alpha$, and

$$\hat{I}_N = \frac{1}{N}\sum_{k=1}^{N}\hat{G}_N(k), \tag{3.12}$$

the test rejects if $\hat{I}_N > c(\alpha, p^2)$ The critical values $c(\alpha, d)$ can be computed using an analytic formula derived by Kiefer [21], but simulating the trajectories of $B_m(\cdot)$ produces critical values which lead to tests with better finite sample properties. The $c(\alpha, d)$ obtained via simulation are tabulated [15] for $d \le 30$.

It is possible to develop a rigorous theory for the behavior of the test under the alternative, but the analysis becomes even more technical and would take up space. We therefore outline only the essential arguments which explain why and when the test is consistent.

First we introduce the following notation: Let $k^* = [n\theta]$, $0 < \theta < 1$, be the time of change. The kernel changes from ψ to ψ^* which satisfies

$$\iint (\psi^*(s,t))^2 \mathrm{d}s\mathrm{d}t < 1.$$

One can show that as $N \to \infty$,

$$\iint (\hat{C}_N(x,y) - \bar{C}(x,y))^2 \mathrm{d}x\mathrm{d}y \xrightarrow{P} 0,$$

where

$$\hat{C}_N(x,y) = \frac{1}{N}\sum_{1\le i\le n} X_i(x)X_i(y)$$

and

$$\bar{C}(x,y) = \theta E[X_0(x)X_0(y)] + (1-\theta)\lim_{N\to\infty} E[X_N(x)X_N(y)].$$

The kernel $\bar{C}(x,y)$ is symmetric, positive-definite and Hilbert–Schmidt with eigenvalues and eigenfunctions $\bar{\lambda}_i$ and $\bar{v}_i$. One can show that as $N \to \infty$, $\|\hat{v}_i - \bar{v}_i\|$ and $|\hat{\lambda}_i - \bar{\lambda}_i|$ tend to 0 in probability.

An application of the ergodic theorem yields that for all $0 \le u \le \theta$,

$$\frac{1}{N}\sum_{1\le i\le Nu}\langle X_{i-1}, \hat{v}_j\rangle\langle X_i, \hat{v}_\ell\rangle \to u\iint R(t,s)\bar{v}_j(t)\bar{v}_\ell(s)\mathrm{d}t\mathrm{d}s \quad \text{a.s.},$$

where $R(t,s) = E[X_1(t)X_2(s)]$.

Under the alternative, $X_{k^*+1}, X_{k^*+2}, \ldots, X_N, X_{N+1}, \ldots$ is not stationary (X_{k^*} is not the stationary initial value), but because $\iint(\psi^*(s,t))^2\mathrm{d}s\mathrm{d}t < 1$ the effect of X_{k^*} is dying out exponentially fast and the elements of X_{k^*+m} are very close to a stationary solution if m is large. So carefully applying the ergodic theorem again, we obtain for all $\theta \le u \le 1$,

$$\frac{1}{N}\sum_{Nu\le i\le N}\langle X_{i-1}, \hat{v}_j\rangle\langle X_i, \hat{v}_\ell\rangle \xrightarrow{P} (1-u)\iint R^*(t,s)\bar{v}_j(t)\bar{v}_\ell(s)\mathrm{d}t\mathrm{d}s,$$

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

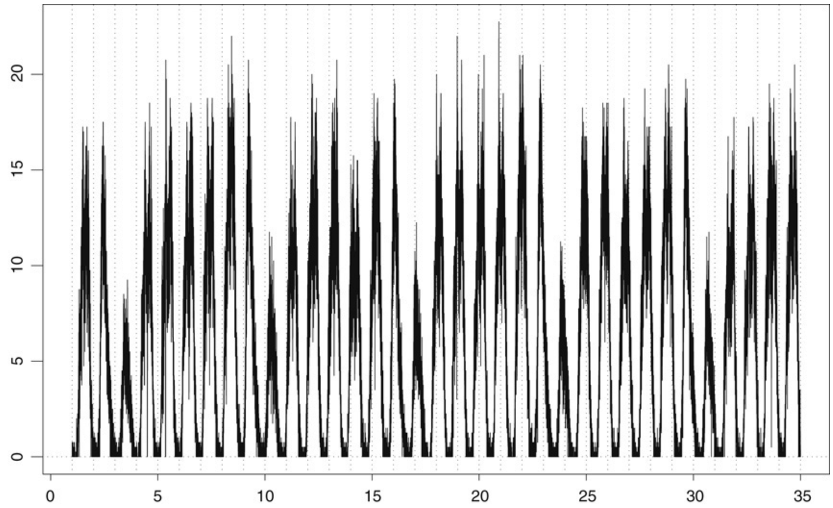

Fig. 2. Five weeks of the original time series of Y_n. Vertical dotted line separate days.

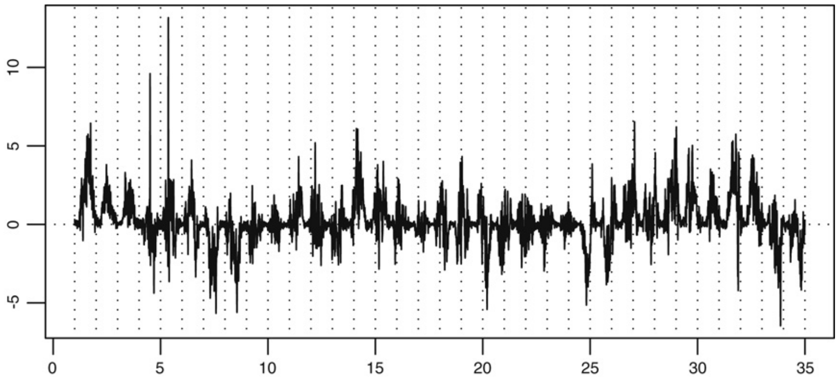

Fig. 3. Five weeks of the differenced time series of X_n.

where $R^*(t,s) = \lim_{N\to\infty} EX_N(t)X_{N+1}(s)$. This means that we have consistency if for at least one (j, ℓ)

$$\iint R(t,s)\bar{v}_j(t)\bar{v}_\ell(s)\mathrm{d}t\mathrm{d}s \neq \int R^*(t,s)\bar{v}_j(t)\bar{v}_\ell(s)\mathrm{d}t\mathrm{d}s,$$

i.e. if R and R^* are different on the space spanned by $\{\bar{v}_j(t)\bar{v}_\ell(s),\ 1 \le j, \ell \le p\}$.

We conclude this section with a summary of the practical implementation of the test procedure:

(1) Find p so large that $\sum_{j=1}^{p} \hat{\lambda}_j / \sum_{j=1}^{N} \hat{\lambda}_j > 0.9$, but not greater than 5.

(2) Compute $\hat{I}_N$ (3.12).

(3) Choose a significance level α and find the critical value $c(\alpha, d)$ with $d = p^2$ from the table in [15].

(4) Reject H_0 if $\hat{I}_N > c(\alpha, p^2)$.

In step (1), p cannot be too large because it is then difficult to estimate $\boldsymbol{D}$. In step (2) good results are obtained if in (3.10) $\frac{k}{N}\widehat{\boldsymbol{D}}_k + \left(1 - \frac{k}{N}\right)\widehat{\boldsymbol{D}}^*_{N-k}$ is replaced by $\widehat{\boldsymbol{D}}_N$, the computations are then much faster.

4. Application to credit card transactions data

In this section we report the results of a small simulation study that examined the finite sample performance of our test. Calculations were performed using the R package `fda`. We worked with a data set studied in [9] which consist of detailed records of transactions made with credit cards issued by Vilnius Bank, Lithuania. The functional time series we study is the count $Y_n(t)$ of transactions in a one minute interval starting at minute t on day $n = 1, 2, \ldots, 200$. The first 35 functional observations are displayed in Fig. 2. To remove weekly seasonality and nonzero mean, the data was differenced at (functional) lag 7, to give $X_n(t) = Y_{n+7}(t) - Y_n(t)$, $n = 1, \ldots, 193$, shown in Fig. 3. The functional data X_n are used in the following. A characteristic pattern of an AR(1) process with clusters of positive and negative observations is clearly seen.

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

Table 1
Empirical size (in percent).

	$p=2$			$p=3$			$p=4$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
$N=50$	9.4	3.4	0.3	11.9	5.9	0.4	6.2	1.9	0.0
$N=100$	9.7	3.6	0.6	9.9	5.0	1.0	7.2	2.3	0.5
$N=200$	8.1	3.8	0.5	10.3	4.8	0.8	6.3	2.8	0.3

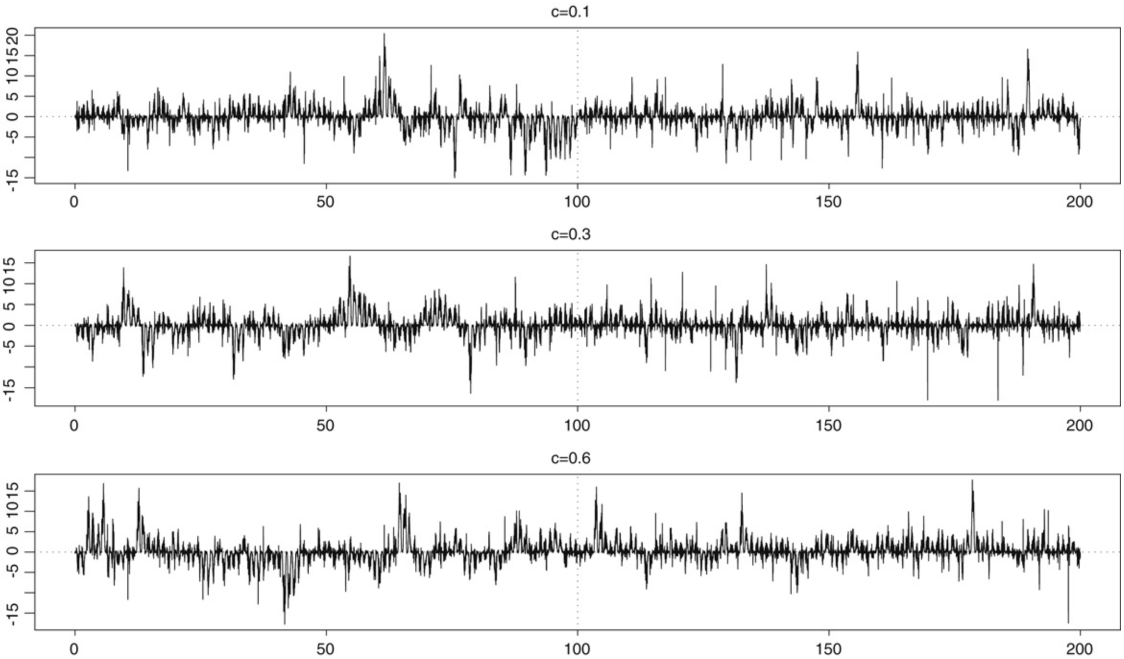

Fig. 4. Bootstrap realizations under alternatives.

Table 2
Empirical power (in percent) for a change occurring at $k^*=N/2$, and $\hat{\psi}$ changing to $c\hat{\psi}$ for $c=0.3$ (in parentheses $c=0.6$).

	$p=2$			$p=3$		
	10%	5%	1%	10%	5%	1%
$N=50$	46.1 (30.9)	28.3 (16.5)	6.3 (1.7)	23.1 (15.9)	10.4 (5.6)	0.3 (0.1)
$N=100$	82.5 (58.1)	67.7 (44.3)	33.5 (16.7)	64.4 (46.9)	46.9 (28.8)	18.2 (7.8)
$N=200$	98.7 (91.6)	95.8 (81.6)	82.3 (52.8)	96.3 (82.8)	90.4 (67.4)	65.6 (34.9)

Applied to these data, our test does not reject the null hypothesis, indicating that a functional AR(1) model is appropriate for the X_n. This is in agreement with the conclusions of Laukaitis and Rackauskas [9] and Gabrys and Kokoszka [28].

In the following, we use the curves X_n to generate functional AR(1) processes which will allow us to assess the finite sample performance of our test in a realistic setting. To do it, we estimate the kernel $\psi(\cdot,\cdot)$ using the function `linmod`, see [29] (we omit the details of regularization). Then, residual functions are computed as $\hat{\varepsilon}_n(t)=X_{n+1}(t)-\hat{\Psi}X_{n+1}(t)$, $n=1,\ldots,193$. Drawing these residuals with replacement, we can simulate functional AR(1) series of any length via

$$Z_m(t)=\int\hat{\psi}(t,s)Z_{m-1}(s)\mathrm{d}s+\varepsilon_m^*(t),\quad m=1,2,\ldots,N,$$

where the $\varepsilon_m^*(\cdot)$ are the bootstrap draws of the $\hat{\varepsilon}_n(\cdot)$. If we change the kernel $\psi(\cdot,\cdot)$ at some point, we can assess the power of the test. To remove the initialization effect, the first "burn-in" 100 simulated functional observations were removed. The empirical rejection rates reported below are based on one thousand replications. To implement the test we have to estimate the long run covariance matrix $\boldsymbol{D}$ (3.3). We used the code of Hansen [30] with some modifications.

Table 1 shows empirical sizes for several values of p and N. The test becomes conservative as p increases. This is because the critical values increase in proportion to p^2, but only the first few principal components explain most of the variance. The same phenomenon was observed in [28,17]. To save space, we report the empirical power only for $p=2$ and $p=3$;

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

for $p = 4$ the power is about 30% lower than for $p = 3$. We introduced a change at half length by multiplying $\hat{\psi}(\cdot,\cdot)$ by $c = 0.1, 0.3, 0.6$, sample realizations for $N = 200$ are shown in Fig. 4. The change is not readily seen by eye, especially for $c = 0.6$. For $c = 0.1$, the second half of the series looks more like white noise, and the power is correspondingly very close to 100%, and so is not reported. Table 2 shows that the power increases with the sample size N, and is satisfactory for $N = 200$, supporting the claim the the ARH(1) model is suitable for the whole credit card transaction record.

5. Asymptotic results

In order to develop an asymptotic theory, we must verify that the test statistic does not change if the principal components $\hat{v}_j$ are replaced by $\hat{c}_j\hat{v}_j$, as only the latter converge to the population principal components v_j. For this purpose, it is convenient to introduce a $p \times p$ diagonal matrix $\boldsymbol{C}_p$ and a $p^2 \times p^2$ diagonal matrix $\boldsymbol{M}$ defined by

$$\boldsymbol{C}_p = \begin{bmatrix} \hat{c}_1 & & & \\ & \hat{c}_2 & & \\ & & \ddots & \\ & & & \hat{c}_p \end{bmatrix}, \qquad \boldsymbol{M} = \boldsymbol{C}_p \otimes \boldsymbol{C}_p,$$

where $\otimes$ denotes the Kronecker product, see e.g. [31]. For example, if $p = 2$

$$\boldsymbol{M} = \begin{bmatrix} \hat{c}_1\hat{c}_1 & & & \\ & \hat{c}_1\hat{c}_2 & & \\ & & \hat{c}_2\hat{c}_1 & \\ & & & \hat{c}_2\hat{c}_2 \end{bmatrix}.$$

Replacing $\hat{v}_j$ by $\hat{c}_j\hat{v}_j$ implies replacing the vectors $\hat{\mathbf{Y}}_i$ by $\boldsymbol{M}\hat{\mathbf{Y}}_i$, which in turn implies replacing $\hat{\mathbf{U}}_N(k)$ by $\boldsymbol{M}\hat{\mathbf{U}}_N(k)$, while $\widehat{\boldsymbol{D}}_k$ and $\widehat{\boldsymbol{D}}^*_{N-k}$ are replaced, respectively, by $\boldsymbol{M}\widehat{\boldsymbol{D}}_k\boldsymbol{M}^{\mathrm{T}}$ and $\boldsymbol{M}\widehat{\boldsymbol{D}}^*_{N-k}\boldsymbol{M}^{\mathrm{T}}$. Since $\boldsymbol{M}^2$ is a $p^2 \times p^2$ identity matrix, it follows that the $\hat{G}(k)$ (3.10) are invariant to the signs of the $\hat{v}_j$. To develop asymptotic arguments, we can thus work with quantities $\hat{c}_j\langle X_i, \hat{v}_j\rangle$ in place of the actual scores $\langle X_i, \hat{v}_j\rangle$.

Recall the definition (3.10) of $\hat{G}(k)$ and introduce the process

$$\hat{Q}_N(u) = \hat{G}_N([Nu]), \quad u \in [0, 1].$$

Recall also the definition of the Bartlett estimators (3.8) and (3.9), and introduce the following assumption of the rate of growth of the bandwidth $q = q(N)$.

Assumption 5.1. Suppose $q(N)$ is nondecreasing and satisfies

$$\sup_{k\ge 0} \frac{q(2^{k+1})}{q(2^k)} < \infty \tag{5.1}$$

and

$$q(N) \to \infty \quad \text{and} \quad q(N)(\log N)^4 = O(N). \tag{5.2}$$

The following theorem shows that the test procedure described in Section 3 has asymptotically correct size.

Theorem 5.1. *Under Assumptions 2.1, 2.2 and 5.1,*

$$\hat{Q}_N(u) \to \sum_{1\le m\le p^2} B_m^2(u) \quad \text{in } D([0,1]),$$

where $\{B_m(u),\ u \in [0,1]\},\ 1 \le m \le p^2$, are iid Brownian bridges.

As we discussed in the previous section, the proof of Theorem 5.1 is split into two steps. The first step, Proposition 5.1 is the weak convergence of the process $Q_N(u) = G_N([Nu]),\ u \in [0,1]$, where G_N is defined by (3.7). This is the CUSUM process based on the projections on population eigenfunctions of the covariance operator. In the second step, Proposition 5.2, it is shown that the estimation of the eigenfunctions and eigenvalues has only asymptotically negligible effect. The second step is more delicate, relies on the special structure of the process Q_N, a truncation and blocking technique, and an application of Mensov's inequality.

Proposition 5.1. *Under Assumption 2.1,*

$$Q_N(u) \to \sum_{1\le m\le p^2} B_m^2(u) \quad \text{in } D([0,1]),$$

where $Q_N(u) = G_N([Nu]),\ u \in [0,1]$, and G_N is defined by (3.7).

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

Proposition 5.2. *Under Assumptions 2.1 and 2.2,*

$$N^{-1/2} \max_{2\le k\le N} \left\| \boldsymbol{M}\hat{\mathbf{U}}_N(k) - \mathbf{U}_N(k) \right\| \xrightarrow{P} 0.$$

Propositions 5.1 and 5.2 are proven, respectively, in Sections 6 and 7. Using them, it is easy to prove Theorem 5.1.

Proof of Theorem 5.1. Recall that $Q_N(u) = G_N([Nu])$, $u \in [0, 1]$, where G_N is defined by (3.7). By Proposition 5.1, $Q_N(u) \to \sum_{1\le m\le p^2} B_m^2(u)$ in $D([0, 1])$. To complete the proof, we must show that

$$\max_{2\le k\le N} |\hat{G}_N(k) - G_N(k)| \xrightarrow{P} 0. \tag{5.3}$$

Relation (5.3) will follow once we have verified that

$$N^{-1/2} \max_{2\le k\le N} \left\| \boldsymbol{M}\hat{\mathbf{U}}_N(k) - \mathbf{U}_N(k) \right\| \xrightarrow{P} 0 \tag{5.4}$$

and

$$\max_{2\le k\le N} \left\| \left[\frac{k}{N}\widehat{\boldsymbol{D}}_k + \left(1 - \frac{k}{N}\right) \widehat{\boldsymbol{D}}^*_{N-k} \right]^{-1} - \boldsymbol{D}^{-1} \right\| \xrightarrow{P} 0. \tag{5.5}$$

Relation (5.4) is stated as Proposition 5.2. To prove (5.5), we use Theorem A.1 and Remark A.1 of [32] which imply that under Assumption 5.1, $\widehat{\boldsymbol{D}}_k$ and $\widehat{\boldsymbol{D}}^*_{N-k}$ converge almost surely to $\boldsymbol{D}$. Recall that if a sequence ζ_n converges to zero a.s., then $\max_{1\le n\le N} |\zeta_n| \xrightarrow{P} 0$, as $N \to \infty$. Therefore, $\sup_{1<u<1} \|u\widehat{\boldsymbol{D}}_{[Nu]} - u\boldsymbol{D}\| \xrightarrow{P} 0$ and $\sup_{1<u<1} \|(1-u)\widehat{\boldsymbol{D}}^*_{N-[Nu]} - (1-u)\boldsymbol{D}\| \xrightarrow{P} 0$, and so

$$\sup_{1<u<1} \|u\widehat{\boldsymbol{D}}_{[Nu]} + (1-u)\widehat{\boldsymbol{D}}^*_{N-[Nu]} - \boldsymbol{D}\| \xrightarrow{P} 0.$$

Since the inverse is a continuous map, (5.5) follows. ■

6. Proof of Proposition 5.1

Proposition 5.1 follows from Proposition 6.1 because by (6.1),

$$Q_N(u) \to [\mathbf{W}_D(u) - u\mathbf{W}_D(1)]^{\mathrm{T}} \boldsymbol{D}^{-1} [\mathbf{W}_D(u) - u\mathbf{W}_D(1)] \quad \text{in } D([0, 1])$$

and a direct computation shows that the Gaussian vectors-valued processes $\{\boldsymbol{D}^{-1/2}[\mathbf{W}_D(u) - u\mathbf{W}(u)],\ u \in [0, 1]\}$ and $\{[B_1(u), \ldots, B_{p^2}(u)]^{\mathrm{T}},\ u \in [0, 1]\}$ have equal covariance functions. Recall that $\mathbf{W}_D(\cdot)$, introduced in Section 3, is a Gaussian process with $E\mathbf{W}_D(u) = 0$ and $E\left[\mathbf{W}_D(u)\mathbf{W}_D^{\mathrm{T}}(s)\right] = \boldsymbol{D} \min(u, s),\ u, s \in [0, 1]$.

Proposition 6.1. *If Assumption 2.1 holds, then*

$$N^{-1/2} \left(\mathbf{Z}_{[Nu]} - E\mathbf{Z}_{[Nu]}\right) \to \mathbf{W}_D(u), \quad \text{in } D^{p^2}([0, 1]). \tag{6.1}$$

Proof. Denote $Z_k(j, \ell) = \sum_{2\le i\le k} Y_i(j, \ell)$. To prove the proposition, it is enough to establish the convergence in $D([0, 1])$ of all linear combinations, namely

$$N^{-1/2} \sum_{j,\ell=1}^{p} \theta(j, \ell) \left(Z_{[Nu]}(j, \ell) - EZ_{[Nu]}(j, \ell)\right) \xrightarrow{d} W_{\theta,D}(u),$$

where $\{W_{\theta,D}(u),\ u \in [0, 1]\}$ is a Brownian motion with variance

$$E\left[W_{\theta,D}^2(u)\right] = u \sum_{j,\ell=1}^{p} \sum_{j',\ell'=1}^{p} \theta(j, \ell)\theta(j', \ell')D(j, \ell; j', \ell').$$

To reduce the notational burden, we focus on just one component, i.e. we want to show that

$$N^{-1/2} \sum_{2\le i\le [Nu]} [Y_i(j, \ell) - EY_i(j, \ell)] \xrightarrow{d} W_{D(i,j)}(u). \tag{6.2}$$

($W_{D(i,j)}(u)$ is defined by setting $\theta(i',j') = \delta_{i'i}\delta_{j'j}$ where $\delta_{..}$ is the Kronecker delta.) Convergence (6.2) (in $D([0,1])$) follows essentially from Theorem 19.1 of [33]; we must verify that the sequence $\{Y_i(j,\ell)\}$ is stationary and ergodic and that

$$\sum_{i=1}^{\infty} |\mathrm{Cov}(Y_0(j,\ell), Y_i(j,\ell))| < \infty. \tag{6.3}$$

Relation (6.3) is established in Lemma 6.1. Ergodicity follows from the representation

$$\begin{aligned} Y_i(j,\ell) &= \left\langle X_{i-1}, v_j \right\rangle [\langle \Psi X_{i-1}, v_\ell\rangle + \langle \varepsilon_i, v_\ell \rangle] \\ &= \left\langle X_{i-1}, v_j \right\rangle \left\langle X_{i-1}, \Psi^{\mathrm{T}} v_\ell \right\rangle + \left\langle X_{i-1}, v_j \right\rangle \left\langle \varepsilon_i, \Psi^{\mathrm{T}} v_\ell \right\rangle \end{aligned}$$

and Theorem 3.1 of [6] (moving average representation of X_k) and Theorem 36.4 of [34] (a function of shifts of an iid sequence forms an ergodic sequence). ■

Now we establish (6.3).

Lemma 6.1. *Under Assumption 2.1, the $Y_i(j,\ell)$ defined by (3.1) satisfy*

$$\sum_{1\le i<\infty} |\mathrm{Cov}(Y_1(j,\ell), Y_i(m,n))| < \infty.$$

Proof. Since

$$\begin{aligned} &Y_i(j,\ell) = \left\langle X_{i-1}, v_j \right\rangle \left\langle X_{i-1}, \Psi^{\mathrm{T}} v_\ell \right\rangle + \left\langle X_{i-1}, v_j \right\rangle \left\langle \varepsilon_i, \Psi^{\mathrm{T}} v_\ell \right\rangle, \\ &\mathrm{Cov}(Y_1(j,\ell), Y_i(m,n)) = C_1(i) + C_2(i) + C_3(i) + C_4(i), \end{aligned}$$

where

$$\begin{aligned} C_1(i) &= \mathrm{Cov}\left(\left\langle X_0, v_j\right\rangle \left\langle X_0, \Psi^{\mathrm{T}} v_\ell\right\rangle, \left\langle X_{i-1}, v_m\right\rangle \left\langle X_{i-1}, \Psi^{\mathrm{T}} v_n\right\rangle\right); \\ C_2(i) &= \mathrm{Cov}\left(\left\langle X_0, v_j\right\rangle \left\langle X_0, \Psi^{\mathrm{T}} v_\ell\right\rangle, \left\langle X_{i-1}, v_m\right\rangle \left\langle \varepsilon_i, \Psi^{\mathrm{T}} v_n\right\rangle\right); \\ C_3(i) &= \mathrm{Cov}\left(\left\langle X_0, v_j\right\rangle \left\langle \varepsilon_1, \Psi^{\mathrm{T}} v_\ell\right\rangle, \left\langle X_{i-1}, v_m\right\rangle \left\langle X_{i-1}, \Psi^{\mathrm{T}} v_n\right\rangle\right); \\ C_4(i) &= \mathrm{Cov}\left(\left\langle X_0, v_j\right\rangle \left\langle \varepsilon_1, \Psi^{\mathrm{T}} v_\ell\right\rangle, \left\langle X_{i-1}, v_m\right\rangle \left\langle \varepsilon_i, \Psi^{\mathrm{T}} v_n\right\rangle\right). \end{aligned}$$

It is easy to see that $C_2(i) = C_4(i) = 0$, for $i > 1$, so it remains to find an absolutely convergent bounds on $C_1(i)$ and $C_3(i)$. We focus on the term $C_1(i)$, the argument for $C_3(i)$ being similar. Consider arbitrary $x, y, u, v \in L^2([0,1])$. Since $X_k = \Psi^k X_0 + \sum_{j=0}^{k-1} \Psi^j \varepsilon_{k-j}$,

$$\mathrm{Cov}\left(\langle X_0, x\rangle \langle X_0, y\rangle, \langle X_k, u\rangle \langle X_k, v\rangle\right) = \mathrm{Cov}\left(\langle X_0, x\rangle \langle X_0, y\rangle, \left\langle \Psi^k X_0, u\right\rangle \left\langle \Psi^k X_0, v\right\rangle\right).$$

Consequently

$$\begin{aligned} &|\mathrm{Cov}\left(\langle X_0, x\rangle \langle X_0, y\rangle, \langle X_k, u\rangle \langle X_k, v\rangle\right)| \\ &\quad \le E\left|\langle X_0, x\rangle \langle X_0, y\rangle \left\langle \Psi^k X_0, u\right\rangle \left\langle \Psi^k X_0, v\right\rangle\right| + E\left|\langle X_0, x\rangle \langle X_0, y\rangle\right| E\left|\langle X_k, u\rangle \langle X_k, v\rangle\right| \\ &\quad \le \|\Psi\|^{2k} \left\{E\|X_0\|^4 + \left[E\|X_0\|^2\right]^2\right\} \|x\|\,\|y\|\,\|u\|\,\|v\|. \end{aligned}$$

Therefore

$$|C_1(i)| \le \|\Psi\|^{2(i-1)} \left\{E\|X_0\|^4 + \left[E\|X_0\|^2\right]^2\right\} \|v_j\|\|v_\ell\|\,\|\Psi^{\mathrm{T}} v_m\|\,\|\Psi^{\mathrm{T}} v_n\| \le 2\|\Psi\|^{2i} E\|X_0\|^4. \quad ■$$

7. Proof of Proposition 5.2

Denote $r(t,s) = E[X_1(t)X_2(s)]$ and

$$\hat{R}(j,\ell) = \iint r(t,s)\hat{u}(t,s)\mathrm{d}t\mathrm{d}s,$$

where

$$\hat{u}(t,s) = v_j(t)v_\ell(s) - \hat{c}_j\hat{v}_j(t)\hat{c}_\ell\hat{v}_\ell(s), \quad 0 < s,t < 1. \tag{7.1}$$

Proof of Proposition 5.2. The component of $\boldsymbol{M}\hat{\mathbf{U}}_N(k) - \mathbf{U}_N(k)$ corresponding to the product of the jth and the ℓth score is equal to

$$\frac{k(N-k)}{N}\left\{\frac{1}{k}\left[\hat{c}_j\hat{c}_\ell\hat{Z}_k(j,\ell) - Z_k(j,\ell) - k\hat{R}(j,\ell)\right] - \frac{1}{N-k}\left[\hat{c}_j\hat{c}_\ell\hat{Z}^*_{N-k}(j,\ell) - Z^*_{N-k}(j,\ell) - (N-k)\hat{R}(j,\ell)\right]\right\}.$$

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

Thus the claim will follow once we have verified that

$$N^{-1/2} \max_{2\le k\le N} \left[\hat{c}_j\hat{c}_\ell\hat{Z}_k(j,\ell) - Z_k(j,\ell) - k\hat{R}(j,\ell)\right] \xrightarrow{P} 0 \tag{7.2}$$

and

$$N^{-1/2} \max_{2\le k\le N} \left[\hat{c}_j\hat{c}_\ell\hat{Z}^*_{N-k}(j,\ell) - Z^*_{N-k}(j,\ell) - (N-k)\hat{R}(j,\ell)\right] \xrightarrow{P} 0. \tag{7.3}$$

Since the above two relations are verified in the same way, we will show only the verification of (7.2).

Observe that

$$Z_k(j,\ell) = \iint \sum_{2\le i\le k} X_{i-1}(t)X_i(s)v_j(t)v_\ell(s)\mathrm{d}t\mathrm{d}s$$

and

$$\hat{c}_j\hat{c}_\ell\hat{Z}_k(j,\ell) = \iint \sum_{2\le i\le k} X_{i-1}(t)X_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_\ell\hat{v}_\ell(s)\mathrm{d}t\mathrm{d}s.$$

Therefore

$$\begin{aligned} Z_k(j,\ell) - \hat{c}_j\hat{c}_\ell\hat{Z}_k(j,\ell) &= \iint \sum_{2\le i\le k} [X_{i-1}(t)X_i(s) - r(t,s)]v_j(t)v_\ell(s)\mathrm{d}t\mathrm{d}s \\ &\quad - \iint \sum_{2\le i\le k} [X_{i-1}(t)X_i(s) - r(t,s)]\hat{c}_j\hat{v}_j(t)\hat{c}_\ell\hat{v}_\ell(s)\mathrm{d}t\mathrm{d}s + (k-1)\int\int r(t,s)[v_j(t)v_\ell(s) - \hat{c}_j\hat{v}_j(t)\hat{c}_\ell\hat{v}_\ell(s)]\mathrm{d}t\mathrm{d}s \\ &= \iint \sum_{2\le i\le k} [X_{i-1}(t)X_i(s) - r(t,s)]\hat{u}(t,s)\mathrm{d}t\mathrm{d}s + (k-1)\hat{R}(j,\ell). \end{aligned}$$

As $\hat{R}(j,\ell) = O_P(1)$, to prove (7.2), it thus remains to show that

$$\max_{2\le k\le N} \left| \iint \sum_{2\le i\le k} [X_{i-1}(t)X_i(s) - r(t,s)]\hat{u}(t,s)\mathrm{d}t\mathrm{d}s \right| = o_P(N^{1/2}). \tag{7.4}$$

Since

$$\begin{aligned} &\iint \left| \sum_{1\le i\le k} [X_{i-1}(t)X_i(s) - r(t,s)] \right| \left|\hat{u}(t,s)\right| \mathrm{d}t\mathrm{d}s \\ &\le \left(\iint \left| \sum_{1\le i\le k} [X_{i-1}(t)X_i(s) - r(t,s)] \right|^2 \mathrm{d}t\mathrm{d}s \right)^{1/2} \left(\iint \left|\hat{u}(t,s)\right|^2 \mathrm{d}t\mathrm{d}s \right)^{1/2}, \end{aligned}$$

(7.4) follows from Lemmas 7.1 and 7.2. ■

Lemma 7.1. *The function $\hat{u} \in L^2([0,1]^2)$ defined by (7.1) satisfies*

$$\|\hat{u}\| = \left(\iint [\hat{u}(t,s)]^2 \mathrm{d}t\mathrm{d}s \right)^{1/2} = O_P(N^{-1/2}).$$

Proof. Since

$$\left|v_j(t)v_\ell(s) - \hat{c}_j\hat{v}_j(t)\hat{c}_\ell\hat{v}_\ell(s)\right|^2 \le 2v_j^2(t)[v_\ell(s) - \hat{c}_j\hat{v}_\ell(s)]^2 + 2\hat{v}_\ell^2(s)[v_j(t) - \hat{c}_\ell\hat{v}_j(t)]^2$$

and v_j and $\hat{v}_\ell$ have unit norm in $L^2([0,1])$, $\|\hat{u}\|^2 \le 2\left\{\|v_\ell - \hat{c}_\ell\hat{v}_\ell\|^2 + \|v_j - \hat{c}_j\hat{v}_j\|^2\right\}$. Consequently, by (2.5), there is a constant K such that $E\|\hat{u}\|^2 \le KN^{-1}$. ■

Lemma 7.2. *Under Assumption 2.1,*

$$N^{-1} \max_{2\le k\le N} \left(\iint \left[\sum_{2\le i\le k} [X_{i-1}(t)X_i(s) - r(t,s)] \right]^2 \mathrm{d}t\mathrm{d}s \right)^{1/2} \xrightarrow{P} 0.$$

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

Proof. By Theorem 3.1 of [6],

$$X_k = \sum_{j=0}^{\infty} \Psi^j \varepsilon_{k-j}, \tag{7.5}$$

where the series converges in the L^2 norm and almost surely. For $c > 0$ to be determined later, introduce the truncated series

$$X_{k,N} = \sum_{j=0}^{c \log N} \Psi^j \varepsilon_{k-j}. \tag{7.6}$$

We will use the decomposition

$$X_{i-1}(t)X_i(s) - r(t,s) = X_{i-1}(t)X_i(s) - X_{i-1,N}(t)X_{i,N}(s) + \left[X_{i-1,N}(t)X_{i,N}(s) - r_N(t,s)\right] + [r_N(t,s) - r(t,s)],$$

where

$$r_N(t,s) = E[X_{i-1,N}(t)X_{i,N}(s)]. \tag{7.7}$$

Introduce also the functions

$$V_{i,N}(t,s) = X_{i-1}(t)X_i(s) - X_{i-1,N}(t)X_{i,N}(s) \tag{7.8}$$

and

$$U_{i,N}(t,s) = X_{i-1,N}(t)X_{i,N}(s). \tag{7.9}$$

To prove the lemma, it suffices to show that

$$N^{-1}E \max_{2\le k\le N} \left\| \sum_{2\le i\le k} V_{i,N} \right\| \to 0, \tag{7.10}$$

$$N^{-1}E \max_{2\le k\le N} \left\| \sum_{2\le i\le k} [U_{i,N} - r_N] \right\| \to 0 \tag{7.11}$$

and

$$\|r_N - r\| \to 0. \tag{7.12}$$

In (7.10), (7.11), (7.12), the norm is taken in the space $L^2([0,1]^2)$.

By Lemma 7.3,

$$E \max_{2\le k\le N} \left\| \sum_{2\le i\le k} V_{i,N} \right\| \le KN^{2-\kappa}, \tag{7.13}$$

for some K and any $\kappa > 0$, provided c is sufficiently large, so (7.10) follows. Relations (7.11) and (7.12) follow, respectively, from Lemmas 7.4 and 7.5. ■

Lemma 7.3. *For $c > 0$ define $X_{k,N} = X_{k,N,c}$ by (7.6). Consider the function $V_{i,N}(t,s)$ defined by (7.8). Then for any $\kappa > 0$, there is c so large that*

$$E\|V_{i,N}\| = E\left\{ \iint V_{i,N}^2(t,s)\mathrm{d}t\mathrm{d}s \right\}^{1/2} \le KN^{-\kappa}$$

for some constant K.

Proof. Observe that

$$\begin{aligned}
\|V_{i,N}\|^2 &= \iint [X_{i-1}(t)X_i(s) - X_{i-1,N}(t)X_{i,N}(s)]^2 \mathrm{d}t\mathrm{d}s \\
&= \iint [X_{i-1}(t)(X_i(s) - X_{i,N}(s)) + X_{i,N}(s)(X_{i-1}(t) - X_{i-1,N}(t))]^2 \mathrm{d}t\mathrm{d}s \\
&\le 2\left\{ \int X_{i-1}^2(t)\mathrm{d}t \int (X_i(s) - X_{i,N}(s))^2 \mathrm{d}s + \int X_{i,N}^2(s)\mathrm{d}s \int (X_{i-1}(t) - X_{i-1,N}(t))^2 \mathrm{d}t \right\} \\
&= 2\left\{ \|X_{i-1}\|^2 \|X_i - X_{i,N}\|^2 + \|X_{i,N}\|^2 \|X_{i-1} - X_{i-1,N}\|^2 \right\}
\end{aligned}$$

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

Define $r > 0$ by $\|\Psi\| = \mathrm{e}^{-r}$. Then

$$E\|X_k - X_{k,N}\| \le \sum_{j > c\log N} \|\Psi\|^j E\|\varepsilon_0\| \le (1-\mathrm{e}^{-r})^{-1} N^{-cr} E\|\varepsilon_0\|,$$

and the claim follows. ■

Lemma 7.4. *The functions $U_{i,N} \in L^2([0,1]^2)$ defined by (7.9) satisfy*

$$E \max_{2\le k\le N} \left\| \sum_{2\le i\le k} \left[U_{i,N} - EU_{i,N}\right] \right\| \le K N^{1/2} (\log N)^{3/2},$$

where K is a constant and the norm is in the space $L^2([0,1]^2)$.

Proof. Set

$$U^*_{i,N}(t,s) = U_{i,N}(t,s) - EU_{i,N}(t,s).$$

Let $m = c\log N$ and assume without loss of generality that m is an integer. We will work with the decomposition

$$\sum_{1\le i\le k} U^*_{i,N} = S_1(k) + S_2(k) + \cdots + S_m(k).$$

The idea is that $S_1(k)$ is the sum of (available) $U^*_{1,N}, U^*_{1+m,N}, \ldots, S_2(k)$ of $U^*_{2,N}, U^*_{2+m,N}, \ldots$, etc. Formally, for $1 \le k \le N$ and $1 \le j \le m$, define

$$S_j(k) = \begin{cases} \displaystyle\sum_{\ell=1}^{[k/m]} U^*_{(\ell-1)m+j,N} + U^*_{m[k/m]+j,N}, & \text{if } k/m \text{ is not an integer} \\ \displaystyle\sum_{\ell=1}^{k/m} U^*_{(\ell-1)m+j,N}, & \text{if } k/m \text{ is an integer.} \end{cases} \tag{7.14}$$

By (7.6) and (7.9), for any fixed j, $S_j(k)$ is a sum of independent identically distributed random functions in $L^2([0,1]^2)$. Since $\left\|\sum_{1\le i\le k} U^*_{i,N}\right\| \le \sum_{j=1}^m \left\|S_j(k)\right\|$,

$$\left\| \sum_{1\le i\le k} U^*_{i,N} \right\|^2 \le m \sum_{j=1}^m \left\|S_j(k)\right\|^2. \tag{7.15}$$

By (7.15) and Lemma 7.6, we obtain $E\left\|\sum_{1\le i\le k} U^*_{i,N}\right\|^2 \le Cmk$, where C is a constant which does not depend on N. Since $U^*_{i,N}$ is a stationary sequence, this bound implies that for all $K < L$,

$$E\left\| \sum_{K\le i\le L} U^*_{i,N} \right\|^2 \le Cm(L-K). \tag{7.16}$$

Relation (7.16) together with the Mensov inequality (Lemma 7.7) imply that

$$E \max_{1\le k\le N} \left\| \sum_{1\le i\le k} U^*_{i,N} \right\|^2 \le Cm(\log N)^2 N. \tag{7.17}$$

Recall that $m = O(\log N)$, to obtain the claim of the lemma. ■

Lemma 7.5. *Recall the functions $r(t,s) = E[X_{i-1}(t)X_i(s)]$ and $r_N(t,s)$ (7.7). Then*

$$\|r - r_N\|^2 = \iint |r(t,s) - r_N(t,s)|^2 \,\mathrm{d}t\mathrm{d}s = O(N^{-2rc}),$$

where $r > 0$ is defined by $\|\Psi\| = \mathrm{e}^{-r}$.

Proof. For ease of notation set $m = c\log N$ and observe that

$$r_N(t,s) = E\left[\sum_{j=0}^m \Psi^j \varepsilon_{i-1-j}(t) \sum_{\ell=0}^m \Psi^\ell \varepsilon_{i-\ell}(t)\right] = \sum_{j=0}^m E\left[\Psi^j \varepsilon_{i-1-j}(t) \Psi^{j+1} \varepsilon_{i-1-j}(t)\right].$$

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

Using an analogous expansion of $r(t,s)$, we obtain

$$\begin{aligned}
\|r - r_N\|^2 &= \iint \left| \sum_{j>m} E[\Psi^j \varepsilon_{-j}(t)\Psi^{j+1}\varepsilon_{-j}(s)] \right|^2 \mathrm{d}t\mathrm{d}s \\
&= \sum_{j,\ell>m} E \iint [\Psi^j \varepsilon_{-j}(t)\Psi^{j+1}\varepsilon_{-j}(s)\Psi^{\ell}\varepsilon_{-l}(t)\Psi^{\ell+1}\varepsilon_{-\ell}(s)]\mathrm{d}t\mathrm{d}s \\
&= \sum_{j,\ell>m} E\left[\int \Psi^j \varepsilon_{-j}(t)\Psi^l \varepsilon_{-\ell}(t)\mathrm{d}t \int \Psi^{j+1}\varepsilon_{-j}(s)\Psi^{\ell+1}\varepsilon_{-\ell}(s)\mathrm{d}s \right] \\
&\le \sum_{j,\ell>m} E\left[\|\Psi^j \varepsilon_{-j}\| \, \|\Psi^{\ell}\varepsilon_{-\ell}\| \, \|\Psi^{j+1}\varepsilon_{-j}\| \, \|\Psi^{\ell+1}\varepsilon_{-\ell}\| \right] \\
&\le \sum_{j,\ell>m} \|\Psi\|^{2j+1}\|\Psi\|^{2\ell+1} E\|\varepsilon_0\|^4 \le K\|\Psi\|^{4m}. \quad \blacksquare
\end{aligned}$$

The following two lemmas are used in the proof of Lemma 7.4.

Lemma 7.6. *The functions $S_j(k) \in L^2([0,1]^2)$ defined by (7.14) satisfy*

$$E\|S_j(k)\|^2 \le Ck/m, \quad 1 \le j \le m,$$

where C is a constant which does not depend on N.

Proof. To lighten the notation, suppose $k/m = n$ is an integer. By stationarity of the $X_{i,N}$,

$$E\|S_j(k)\|^2 = E \iint \left| \sum_{\ell=1}^{n} U^*_{(\ell-1)m+j,N}(t,s) \right|^2 \mathrm{d}t\mathrm{d}s = E \iint \left| \sum_{\ell=1}^{n} U^*_{(\ell-1)m,N}(t,s) \right|^2 \mathrm{d}t\mathrm{d}s$$

does not depend on j. By construction, the $U^*_{(\ell-1)m,N}(t,s)$ are mean zero and $U^*_{(\ell-1)m,N}(t,s)$ is independent of $U^*_{(\ell'-1)m,N}(t,s)$ if $l' \ne \ell$. Therefore

$$E\|S_j(k)\|^2 = \iint \sum_{\ell=1}^{n} E\left[U^{*2}_{(\ell-1)m,N}(t,s)\right]\mathrm{d}t\mathrm{d}s = n \iint E\left[U^{*2}_{1,N}(t,s)\right]\mathrm{d}t\mathrm{d}s.$$

It thus remains to show that $\iint E\left[U^{*2}_{1,N}(t,s)\right]\mathrm{d}t\mathrm{d}s$ is bounded by a constant which does not depend on N.

Observe that

$$\begin{aligned}
\iint E\left[U^{*2}_{1,N}(t,s)\right]\mathrm{d}t\mathrm{d}s &\le \iint E\left[X^2_{0,N}(t)X^2_{1,N}(s)\right]\mathrm{d}t\mathrm{d}s \\
&= E\left(\int X^2_{0,N}(t)\mathrm{d}t\right)\left(\int X^2_{1,N}(s)\mathrm{d}s\right) \le E\left(\int X^2_{0,N}(t)\mathrm{d}t\right)^2 = E\|X_{0,N}\|^4.
\end{aligned}$$

Setting $m = c\log N$, we get

$$\begin{aligned}
E\|X_{0,N}\|^4 &= E\left\| \sum_{j=0}^{m} \Psi^j \varepsilon_{-j} \right\|^4 \le E\left(\sum_{j=0}^{m} \|\Psi\|^j \|\varepsilon_{-j}\| \right)^4 \\
&= \sum_{j_1=0}^{m}\sum_{j_2=0}^{m}\sum_{j_3=0}^{m}\sum_{j_4=0}^{m} \|\Psi\|^{j_1}\|\Psi\|^{j_2}\|\Psi\|^{j_3}\|\Psi\|^{j_4} E\left[\|\varepsilon_{-j_1}\| \, \|\varepsilon_{-j_2}\| \, \|\varepsilon_{-j_3}\| \, \|\varepsilon_{-j_4}\|\right] \\
&\le \left(\sum_{j=0}^{m} \|\Psi\|^j \right)^4 E\|\varepsilon_0\|^4 \le (1 - \|\Psi\|)^{-4} E\|\varepsilon_0\|^4. \quad \blacksquare
\end{aligned}$$

Lemma 7.7 (*Mensov Inequality*). *Let $\xi_1, \xi_2, \ldots$ be arbitrary Hilbert space valued random variables. If for any $K < L$*

$$E\left\| \sum_{i=K+1}^{L} \xi_i \right\|^2 \le C(L-K) \tag{7.18}$$

Please cite this article in press as: L. Horváth, et al., Testing the stability of the functional autoregressive process, Journal of Multivariate Analysis (2009), doi:10.1016/j.jmva.2008.12.008

then, for any b,

$$E \max_{1\leq k\leq N} \left\| \sum_{i=1+b}^{k+b} \xi_i \right\|^2 \leq C[\log(2N)]^2 N. \tag{7.19}$$

Proof. The proof is practically the same as for real-valued random variables ξ_i, see [35], and so is omitted. ■

Acknowledgments

We thank Robertas Gabrys for performing the numerical work reported in Section 4. The first author was partially supported by NSF grant DMS-0604670. The second author was partially supported by grants GAOR 201/06/0186, LC06024 and MSM 0021620839. The third author was partially supported by NSF grants DMS-0413653 and DMS-0804165.

References

[1] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, Springer, 2005.
[2] H.-G. Müller, U. Stadtmüller, Generalized functional linear models, The Annals of Statistics 33 (2005) 774–805.
[3] F. Yao, H.-G. Müller, J.-L. Wang, Functional linear regression analysis for longitudinal data, The Annals of Statistics 33 (2005) 2873–2903.
[4] R.H. Glendinning, S.L. Fleet, Classifying functional time series, Signal Processing 87 (2007) 79–100.
[5] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer, 2006.
[6] D. Bosq, Linear Processes in Function Spaces, Springer, New York, 2000.
[7] P. Besse, H. Cardot, Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1, Canadian Journal of Statistics 24 (1996) 467–487.
[8] P. Besse, H. Cardot, D. Stephenson, Autoregressive forecasting of some functional climatic variations, Scandinavian Journal of Statistics 27 (2000) 673–687.
[9] A. Laukaitis, A. Račkauskas, Functional data analysis of payment systems, Nonlinear Analysis: Modeling and Control 7 (2002) 53–68.
[10] A. Antoniadis, T. Sapatinas, Wavelet methods for continuous time prediction using Hilbert-valued autoregressive processes, Journal of Multivariate Analysis 87 (2003) 133–158.
[11] B. Fernández de Castro, S. Guillas, W. Gonzáles Manteiga, Functional samples and bootstrap for predicting sulfur dioxide levels, Technometrics 47 (2005) 212–222.
[12] V. Kargin, A. Onatski, Curve forecasting by functional autoregression, Journal of Multivariate Analysis 00 (2008) 000–000.
[13] P. Hall, M. Hosseini-Nasab, On properties of functional principal components, Journal of Royal Statistical Society. Series B 68 (2006) 109–126.
[14] M. Benko, W. Härdle, A. Kneip, Common functional principal components, The Annals of Statistics 00 (0000–0000) (2008) (in press).
[15] I. Berkes, R. Gabrys, L. Horváth, P. Kokoszka, Detecting changes in the mean of functional observations, Technical Report, Utah State University, 2007.
[16] A. Laukaitis, On some applications of functional data analysis in econometrics, Ph.D. Thesis, Vilnius University, 2002.
[17] P. Kokoszka, I. Maslova, J. Sojka, L. Zhu, Testing for lack of dependence in functional linear model, Canadian Journal of Statistics 36 (2008) 207–222.
[18] P. Kokoszka, I. Maslova, J. Sojka, L. Zhu, Effect of substorms on the magnetic field variability at mid- and low-latitudes, Technical Report, Utah State University, 2008.
[19] M. Csörgő, L. Horváth, Weighted Approximations in Probability and Statistics, Wiley, New York, 1993.
[20] M. Csörgő, L. Horváth, Limit Theorems in Change-Point Analysis, Wiley, New York, 1997.
[21] J. Kiefer, K-sample analogues of the Kolmogorov–Smirnov and Cramér–v.Mises tests, Annals of Mathematical Statistics 30 (1959) 420–447.
[22] L. Horváth, P.S. Kokoszka, J. Steinebach, Testing for changes in multivariate dependent observations with applications to temperature changes, Journal of Multivariate Analysis 68 (1999) 96–119.
[23] D.W.K. Andrews, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59 (1991) 817–858.
[24] D.W.K. Andrews, J.C. Monahan, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, Econometrica 60 (1992) 953–966.
[25] P.M. Robinson, Inference without smoothing in the presence of nonparametric autocorrelation, Econometrica 66 (1998) 1163–1182.
[26] J. Antoch, M. Hušková, Z. Prášková, Effect of dependence on statistics for determination of change, Journal of Statistical Planning and Inference 60 (1997) 291–310.
[27] M. Hušková, Z. Prášková, J. Steinebach, On the detection of changes in autoregressive time series I. Asymptotics, Journal of Statistical Planning and Inference 137 (2007) 1243–1259.
[28] R. Gabrys, P. Kokoszka, Portmanteau test of independence for functional observations, Journal of the American Statistical Association 102 (2007) 1338–1348.
[29] N. Malfait, J.O. Ramsay, The historical functional model, Canadian Journal of Statistics 31 (2003) 115–128.
[30] B.E. Hansen, Rethinking the univariate approach to unit root testing: Using covariates to increase power, Econometric Theory 11 (1995) 1148–1171. Code available at: http://www.ssc.wisc.edu/~bhansen.
[31] A. Graham, Kronecker Products and Matrix Calculus with Applications, John Wiley and Sons, 1981.
[32] I. Berkes, L. Horváth, P. Kokoszka, Q.-M. Shao, On discriminating between long-range dependence and changes in mean, The Annals of Statistics 34 (2006) 1140–1165.
[33] P. Billingsley, Convergence of Probability Measures, 2nd edition, Wiley, New York, 1999.
[34] P. Billingsley, Probability and Measure, 3rd edition, Wiley, New York, 1995.
[35] F. Móricz, Moment inequalities and the strong law of large numbers, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 35 (1976) 299–314.