ARTICLE IN PRESS
Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Testing for independence between functional time series☆

Lajos Horváth *, Gregory Rice

Department of Mathematics, University of Utah, Salt Lake City, UT 84112-0090, USA

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C12
C32

Keywords:
Functional observations
Tests for independence
Weak dependence
Long run covariance function
Central limit theorem

ABSTRACT

Frequently econometricians are interested in verifying a relationship between two or more time series. Such analysis is typically carried out by causality and/or independence tests which have been well studied when the data is univariate or multivariate. Modern data though is increasingly of a high dimensional or functional nature for which finite dimensional methods are not suitable. In the present paper we develop methodology to check the assumption that data obtained from two functional time series are independent. Our procedure is based on the norms of empirical cross covariance operators and is asymptotically validated when the underlying populations are assumed to be in a class of weakly dependent random functions which include the functional ARMA, ARCH and GARCH processes.

1. Introduction and results

A common goal of data analysis in econometrics is to determine whether or not a relationship exists between two variables which are measured over time. A determination in either way may be useful. On one hand, if a relationship is confirmed to exist between two variables then further investigation into the strength and nature of the relationship may lead to interesting insights or effective predictive models. Conversely if there is no relationship between the two variables then an entire toolbox of statistical techniques developed to analyze two samples that are independent may be used. The problem of testing for correlation between two univariate or multivariate time series has been well treated, and we discuss the relevant literature below. However, as a by product of seemingly insatiable modern data storage technology, many data of interest exhibit such large dimension that traditional multivariate techniques are not suitable. For example, tick by tick stock return data is stored hundreds of times per second, leading to thousands of observations during a single day. In such cases a pragmatic approach is to treat the data as densely observed measurements from an underlying curve, and, after using the measurements to approximate the curve, apply statistical techniques to the curves themselves. This approach is fundamental in functional data analysis, and in recent years much effort has been put forth to adapt currently available procedures in multivariate analysis to functional data. The goal of the present paper is to develop a test for independence between two functional time series.

In the context of checking for independence between two second order stationary univariate time series, Haugh (1976) proposed a testing procedure based on sample cross-correlation estimators. His test may be considered as an adaptation of the popular Box–Ljung–Pierce portmanteau test (cf. Ljung and Box (1978)) to two samples. In a similar progression the multivariate portmanteau test of Li and McLeod (1981) was extended to test for correlation between two multivariate ARMA time series by El Himdi and Roy (1997) whose test statistic was based on cross-correlation matrices. The literature on such tests has also grown over the years to include adaptations for robustness as well as several other considerations, see Koch and Yang (1986), Li and Hui (1994) and El Himdi et al. (2003) for details. Many of these results are summarized in Li (2004). A separate approach for multivariate data based on the distance correlation measure is developed in Székely and Rizzo (2013).

The analysis of functional time series has seen increased attention in statistics, economics and in other applications over the last decade, see Horváth and Kokoszka (2012) for a summary of recent advances. To test for independence within a single functional time series, Gabrys and Kokoszka (2007) proposed a method where the functional observations are projected onto finitely many basis elements, and a multivariate portmanteau test is applied to the vectors of scores which define the projections. Horváth et al. (2013) developed a portmanteau test for functional data in which the inference is performed using the operator norm of the empirical covariance operators at lags h, $1 \leq h \leq H$, which could be considered as a direct functional analog of the Box–Ljung–Pierce test. Due to

☆ Research supported by National Science Foundation grant DMS 1305858.
* Corresponding author.
E-mail address: horvath@math.utah.edu (L. Horváth).

http://dx.doi.org/10.1016/j.jeconom.2015.03.030
0304-4076/

the infinite dimension of functional data, a normal limit is established for the test statistic rather than the classical χ^2 limit with degrees of freedom depending on the data dimension.

The method that we propose for testing noncorrelation between two functional time series follows the example of Horváth et al. (2013). Suppose that we have observed $X_1(t), \ldots, X_n(t)$ and $Y_1(s), \ldots, Y_n(s)$ which are samples from jointly stationary sequences of square integrable random functions on [0, 1]. Formally we are interested in testing

$$H_0: \text{the sequences } \{X_i\}_{i=1}^{\infty} \quad \text{and} \quad \{Y_j\}_{j=1}^{\infty} \text{ are independent}$$

against the alternative

$$H_A: \text{for some integer } h_0, -\infty < h_0 < \infty, \iint C_{h_0}^2(t,s)dtds > 0 \quad \text{where } C_{h_0}(t,s) = \mathrm{Cov}(X_0(t), Y_{h_0}(s)).$$

We use the notation $\int$ to mean $\int_0^1$. Assuming jointly Gaussian distributions for the underlying observations, independence reduces to zero cross-correlations at all lags, and hence H_A is equivalent to the complement of H_0 in that case. To derive the test statistic, we note that under H_0, the sample cross-covariance functions

$$\hat{C}_{n,h}(t,s) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n-h}(X_i(t)-\bar{X}(t))(Y_{i+h}(s)-\bar{Y}(s)) & h \ge 0 \\ \dfrac{1}{n}\displaystyle\sum_{i=1-h}^{n}(X_i(t)-\bar{X}(t))(Y_{i+h}(s)-\bar{Y}(s)) & h < 0 \end{cases}$$

should be close to the zero function for all choices of h, where

$$\bar{X}(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t), \quad \text{and} \quad \bar{Y}(s) = \frac{1}{n}\sum_{i=1}^{n} Y_i(s).$$

Under H_A a cross covariance function is different from the zero function for at least one h. The test statistic is then based on

$$\hat{T}_{n,H} = \sum_{h=-H}^{H} \iint \hat{C}_{n,h}^2(t,s)dtds$$

with the hope that it includes the covariance estimator corresponding to a non zero function if it exists. We then reject H_0 for large values of $\hat{T}_{n,H}$. Our main result is the asymptotic distribution of $\hat{T}_{n,H}$ under H_0.

In order to cover a large class of functional time series processes, we assume that $\mathbf{X} = \{X_i\}_{i=-\infty}^{\infty}$ and $\mathbf{Y} = \{Y_i\}_{i=-\infty}^{\infty}$ are approximable Bernoulli shifts. We say that $\boldsymbol{\eta} = \{\eta_j(t)\}_{j=-\infty}^{\infty}$ is an L^4 absolutely approximable Bernoulli shift in $\{\epsilon_j(t), -\infty < j < \infty\}$ if

$$\eta_i = g(\epsilon_i, \epsilon_{i-1}, \ldots) \text{ for some nonrandom measurable function } g: S^\infty \mapsto L^2 \text{ and i.i.d. random innovations } \epsilon_j, -\infty < j < \infty, \text{ with values in a measurable space } S, \tag{1.1}$$

$$\eta_j(t) = \eta_j(t,\omega) \text{ is jointly measurable in } (t,\omega)\ (-\infty < j < \infty), \tag{1.2}$$

and

$$\text{the sequence } \{\boldsymbol{\eta}\} \text{ can be approximated by } \ell\text{-dependent sequences } \{\eta_{j,\ell}\}_{j=-\infty}^{\infty} \text{ in the sense that } \sum_{\ell=1}^{\infty} \ell (E\|\eta_j - \eta_{j,\ell}\|^4)^{1/4} < \infty \text{ where } \eta_{j,\ell} \text{ is defined by } \eta_{j,\ell} = g(\epsilon_j, \epsilon_{j-1}, \ldots, \epsilon_{j-\ell+1}, \boldsymbol{\epsilon}_{j,\ell}^*), \ \boldsymbol{\epsilon}_{j,\ell}^* = (\epsilon_{j,\ell,j-\ell}^*, \epsilon_{j,\ell,j-\ell-1}^*, \ldots), \text{ where the } \epsilon_{j,\ell,k}^*\text{'s are independent copies of } \epsilon_0, \text{ independent of } \{\epsilon_j, -\infty < j < \infty\}. \tag{1.3}$$

In assumption (1.1) we take S to be an arbitrary measurable space, however in most applications S is itself a function space and the evaluation of $g(\epsilon_i, \epsilon_{i-1}, \ldots)$ is a functional of $\{\epsilon_j(t)\}_{j=-\infty}^{i}$. In this case assumption (1.2) follows from the joint measurability of the $\epsilon_i(t,\omega)$'s. Assumption (1.3) is stronger than the requirement $\sum_{\ell=1}^{\infty}(E\|\eta_j - \eta_{j,\ell}\|^2)^{1/2} < \infty$ used by Hörmann and Kokoszka (2010), Berkes et al. (2013) and Jirak (2013) to establish the central limit theorem for sums of Bernoulli shifts. Since we need the central limit theorem for sample correlations, higher moment conditions and a faster rate in the approximability with ℓ-dependent sequences are needed.

We assume that the sequences $\mathbf{X}$ and $\mathbf{Y}$ satisfy the following conditions:

Assumption 1.1. $E\|X_0\|^{4+\delta} < \infty$ and $E\|Y_0\|^{4+\delta} < \infty$ with some $\delta > 0$,

Assumption 1.2. $\mathbf{X} = \{X_i(t)\}_{i=-\infty}^{\infty}$ is an L^4 absolutely approximable Bernoulli shift in $\{\epsilon_j(t), -\infty < j < \infty\}$,

and

Assumption 1.3. $\mathbf{Y} = \{Y_i(t)\}_{i=-\infty}^{\infty}$ is an L^4 absolutely approximable Bernoulli shift in $\{\bar{\epsilon}_j(t), -\infty < j < \infty\}$.

The functions defining the Bernoulli shift sequences $\mathbf{X}$ and $\mathbf{Y}$ as in (1.1) will be denoted by $g_{\mathbf{X}}$ and $g_{\mathbf{Y}}$, respectively. The independence of the sequences $\mathbf{X}$ and $\mathbf{Y}$ stated under H_0 is conveniently given by:

Assumption 1.4. The sequences $\{\epsilon_j(t), -\infty < j < \infty\}$ and $\{\bar{\epsilon}_j(t), -\infty < j < \infty\}$ are independent.

The parameter H defines the number of lags used to define the test statistic. As n increases it becomes possible to accurately estimate cross covariances for larger lags, and thus we allow H to tend to infinity with the sample size. Namely,

Assumption 1.5. $H = H(n) \to \infty$ and $Hn^{-\tau} \to 0$, as $n \to \infty$, with some $0 < \tau < 2\delta/(4+7\delta)$, where δ is defined in Assumption 1.1.

To state the limit result for $\hat{T}_{n,H}$ we first introduce the asymptotic expected value and variance. Let for all $-\infty < j < \infty$

$$\gamma_{\mathbf{X}}(j) = \int \mathrm{cov}(X_0(t), X_j(t))dt, \qquad \gamma_{\mathbf{Y}}(j) = \int \mathrm{cov}(Y_0(t), Y_j(t))dt$$

and define

$$\mu = \sum_{j=-\infty}^{\infty} \gamma_{\mathbf{X}}(j)\gamma_{\mathbf{Y}}(j). \tag{1.4}$$

It is shown in Lemma B.1 that under Assumptions 1.2 and 1.3 the infinite sum in the definition of μ above is absolutely convergent. Let

$$\sigma_h^2 = 2\int \cdots \int \left(\sum_{\ell=-\infty}^{\infty} \mathrm{cov}(X_0(t), X_\ell(s))\mathrm{cov}(Y_0(u), Y_{\ell+h}(v))\right)^2 \times dtdsdudv$$

and

$$\sigma^2 = \sum_{h=-\infty}^{\infty} \sigma_h^2. \tag{1.5}$$

Theorem 1.1. *If Assumptions 1.1–1.5 hold, then we have*

$$\frac{n\hat{T}_{n,H} - (2H+1)\mu}{(2H+1)^{1/2}\sigma} \xrightarrow{\mathcal{D}} \mathcal{N},$$

where $\mathcal{N}$ stands for a standard normal random variable.

Please cite this article in press as: Horváth, L., Rice, G., Testing for independence between functional time series. Journal of Econometrics (2015), http://dx.doi.org/10.1016/j.jeconom.2015.03.030