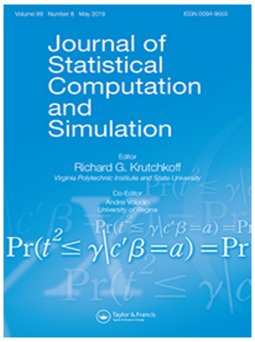

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Improved functional portmanteau tests

Atefeh Zamani, Maryam Hashemi & Hossein Haghbin

To cite this article: Atefeh Zamani, Maryam Hashemi & Hossein Haghbin (2019) Improved functional portmanteau tests, Journal of Statistical Computation and Simulation, 89:8, 1423-1436, DOI: 10.1080/00949655.2019.1584199

To link to this article: https://doi.org/10.1080/00949655.2019.1584199

Published online: 08 Mar 2019.

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
2019, VOL. 89, NO. 8, 1423–1436
https://doi.org/10.1080/00949655.2019.1584199

Improved functional portmanteau tests

Atefeh Zamani[a], Maryam Hashemi[b] and Hossein Haghbin[c]

[a]Department of Statistics, Faculty of Science, Shiraz University, Shiraz, Iran; [b]Statistics Department, University of Khansar, Khansar, Iran; [c]Department of Statistics, Faculty of Science, Persian Gulf University , Boushehr, Iran

ABSTRACT

Functional time series is a popular method of forecasting in functional data analysis. The Box-Jenkins methodology for model building, with the aim of forecasting, includes three iterative steps of model identification, parameter estimation and diagnostic checking. Portmanteau tests are one of the most popular diagnostic checking tools. In particular, they are applied to find if the residuals of the fitted model are white noise. Gabrys and Kokoszka [Portmanteau test of independence for functional observations. J Am Stat Assoc. 2007;102(480):1338–1348.] proposed a portmanteau test of independence for functional observation based on Box and Pierce's statistic. Their statistic is too sensitive to the lag value, specially when the sample size is small. Here, two modifications of Gabrys and Kokoszka statistic are presented, which have superior properties in small samples. The efficiency of the modified statistics is demonstrated through a simulation study.

ARTICLE HISTORY
Received 11 March 2018
Accepted 13 February 2019

KEYWORDS
Functional time series; portmanteau test; independence.

MSC2000 CLASSIFICATIONS
62M10; 62H15

1. Introduction

Time series analysis plays a crucial role in statistical modelling, where the main purpose is forecasting. The Box-Jenkins methodology, which is an appropriate method for a long-term forecasting, is consisted of model identification, parameter estimation, diagnostic checking and forecasting. Generally speaking, the purpose of model identification is to find an appropriate model for the data, such that the residuals can be modelled as a sequence of uncorrelated error terms. When the statistical model is identified, the parameters of the model are estimated and the residuals are obtained. The next step in model development, which will be followed by forecasting, is to test the properties of the residuals. Portmanteau tests, by checking the significance of the residual autocorrelations, are important tools for diagnostic checking of fitted time series models.

One of the most applicable models in the analysis on time series data is autoregressive-moving average models and Box and Pierce [1] are the first researchers who obtained a portmanteau test for goodness of fit test in such processes, using a statistic based on the residuals of the fitted model. In [2] presented a modification of the earliest form of Box and Pierce statistic and the multivariate adjustment of this modification is studied by Hosking [3]. Besides, Li and McLeod [4] tried to improve the performance of the portmanteau

CONTACT Maryam Hashemi hashemi@khansar-cmc.ac.ir

test by adding some term to the classic one. McLeod and Li [5] and Monti [6] presented other forms of univariate portmanteau test statistic. Moreover, the multivariate portmanteau tests are studied by researchers, such as Chitturi [7], Francq and Raïssi [8] and Mahdi and McLeod [9].

Advancing modern technologies result in recording data during a continuous time interval or at several discrete intermittent time points. These cases can be considered as 'functional data', which have become a commonly encountered type of data, through these years. Functional Data Analysis (FDA) is concerned with the statistical methodology for the analysis of such data and various researchers have been trying to extend the classical methods of data analysis in this field. For more on FDA methods and applications, we refer readers to Ramsay and Silverman [10], Ferraty and Vieu [11] and Horváth and Kokoszka [12] and the references given there.

Portmanteau tests in the analysis of functional time series is introduced by Gabrys and Kokoszka [13]. They proposed a simple portmanteau test of independence for functional observations, which is based on Box and Pierce's statistic. They presented basis for residual-based specification test for various functional time series. Our main concern in this research is modifying Gabrys and Kokoszka statistic to improve the test performance, especially for small samples.

The rest of this article is organized as follows. In Section 2, along with a brief review of functional portmanteau test, preliminary notations and definitions are presented. Two modified test statistics and their asymptotic distributions are obtained in Section 3. Section 4 is devoted to the portmanteau tests in functional autoregressive processes of order one and the consistency of the statistics are studied in this model. Section 5 demonstrates improved performance of the proposed statistics compared with Gabrys and Kokoszka statistic by simulation and real data.

2. Preliminaries

Let $L^2[0,1)$ denotes the Hilbert space of all square integrable functions on $[0,1)$, which is endowed with the inner product $\langle f,g\rangle = \int_0^1 f(t)g(t)\mathrm{d}t$ and the norm $\|f\| = (\int_0^1 f^2(t)\mathrm{d}t)^{1/2}$. Let $\mathcal{L}$ stands for the class of bounded linear operators on $L^2[0,1)$. Random functions $\{X_n(t),\ t \in [0,1),\ n = 1,2,\ldots\}$ are measurable elements of $L^2[0,1)$ and, for theoretical convenience and throughout the paper, we assume that X_n, $n = 1,2,\ldots,$ are zero-mean random functions. Note that, in FDA, $X_i(t)$ is commonly defined as the variable X measured at time t for subject i and, in general, $X_i(t)$'s are independent across i. However, in functional time series $X_n(t)$, t stands for the argument of the function and we use n to denote the time index, which demonstrates the dependence among observations.

Let X stands for a random element with an arbitrary distribution. The covariance operator, $C \in \mathcal{L}$, which is a compact, symmetric, self-adjoint and nuclear operator, is defined as

$$C\left(x\right) = \mathbb{E}\left[\left(X \otimes X\right) x\right] = \mathbb{E}\left[\langle X, x\rangle\, X\right], \quad x \in L^2\left[0,1\right),$$

where $\mathbb{E}$ denotes the Bochner integral on $L^2[0,1)$. Besides, let us denoted the eigenfunctions and eigenvalues of C by ν_j and λ_j, $j = 1,2,\ldots,$ respectively. These elements can be calculated as $C(\nu_j) = \lambda_j\nu_j$, $j = 1,2,\ldots$ and we assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Additionally, the eigenfunctions of C form an orthonormal basis for $L^2[0,1)$ and, consequently,

X can be represented as

$$X = \sum_{k=1}^{\infty} X_k \nu_k, \tag{1}$$

where X_k are the random weights or scores corresponding to X:

$$X_k := \langle X, \nu_k \rangle = \int_0^1 X(t)\, \nu_k(t)\, \mathrm{d}t, \quad k = 1, 2, \ldots. \tag{2}$$

In practical problems, the covariance operator is generally unknown. Observing the random functions $\{X_n(t),\ t \in [0,1),\ n = 1, 2, \ldots, N\}$, the empirical covariance operator can be defined as $C_N(x) = (1/N)\sum_{n=1}^{N} \langle X_n, x\rangle X_n$, $x \in L^2[0,1)$. Similarly, the covariance and empirical covariance operators of lag h are defined as:

$$C_h(x) = \mathbb{E}\left[\left(X_{n+h} \otimes X_n\right) x\right] = \mathbb{E}\left[\left\langle X_{n+h}, x\right\rangle X_n\right] \tag{3}$$

$$C_{N,h}(x) = (1/N) \sum_{n=1}^{N-h} \left\langle X_{n+h}, x\right\rangle X_n, \tag{4}$$

respectively.

Let $X_n(t),\ t \in [0,1),\ n = 1, 2, \ldots, N$, be the sequence of residuals of the fitted model in functional time series. The purpose of portmanteau test is to present goodness of fit test based on the $X_n(t)$'s. In fact, we are going to test $H_0 : X_n(\cdot)$ are independent and identically distributed versus $H_1 : H_0$ does not hold. In [13], Gabrys and Kokoszka proposed a portmanteau test for functional observations. Their idea is simple and intuitively appealing since they approximated the functional observations $X_n(t),\ t \in [0,1),\ n = 1, 2, \ldots, N$, by the first p terms of the principal component expansion,

$$X_n(t) \approx \sum_{k=1}^{p} X_{kn} \nu_k(t), \quad n = 1, 2, \ldots, N, \tag{5}$$

where $\nu_k(t)$'s are the principal components (PCs) of the covariance operator, and the X_{kn}'s are defined as in (2). In reality, ν_k's are unknown and should be estimated using the empirical covariance operator, as follows:

$$C_N\left(\hat{\nu}_k\right) = \hat{\lambda}_k \hat{\nu}_k, \quad k = 1, 2, \ldots. \tag{6}$$

By substituting the empirical eigenvalues and empirical eigenfunctions, $\hat{\lambda}_k$ and $\hat{\nu}_k$, $k = 1, 2, \ldots$, in (5), we have:

$$\hat{X}_n(t) \approx \sum_{k=1}^{p} \hat{X}_{kn} \hat{\nu}_k(t), \quad n = 1, 2, \ldots, N. \tag{7}$$

In practical applications, the choice of p is an important issue and directly affects the accuracy of the estimation procedure. Different methods, such as scree plot, marginal BIC, conditional AIC, pesudo-AIC and cross-validation, are applied to determine the value of

p, [12] and Li et al. [14]. However, in practice, the cumulative percentage of total variance (CPV) is more popular and is defined as

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

which depends on the first p empirical functional principal components. The value of p is chosen such that CPV(p) exceeds 0.85, which is the recommended value, Horváth and Kokoszka [12].

In order to determine the distribution of the test statistic, the following assumption is required:

Assumption 2.1: *The functional observations X_n are mean zero with finite fourth moment, i.e.*

$$E\,\|X_n\|^4 = E\left[\int_0^1 X_n^2\,(t)\;\mathrm{d}t\right]^2 < \infty,$$

and, for $k \leq p$, $\limsup_{N\to\infty} NE\|\nu_k - \hat{\nu}_k\|^2 < \infty$.

Let $r_{f,h}(i,j)$ and $r_{b,h}(i,j)$ denote the (i,j) entries of $\boldsymbol{C}_0^{-1}\boldsymbol{C}_h$ and $\boldsymbol{C}_h\boldsymbol{C}_0^{-1}$, where $\boldsymbol{C}_h$ denotes the sample autocovariance matrix with entries $c_h(k,l)$:

$$c_h\,(k,l) = \frac{1}{N}\sum_{t=1}^{N-h} X_{kt}X_{l,t+h}, \quad 0 \leq h < N. \tag{8}$$

The following portmanteau test statistic, Q_N, is introduced by Gabrys and Kokoszka [13] for functional time series:

$$Q_N = N\sum_{h=1}^{H}\sum_{i,j=1}^{p} r_{f,h}\,(i,j)\;\; r_{b,h}\,(i,j)\,. \tag{9}$$

They also provide its empirical counterpart, $\hat{Q}_N$, by substituting X_n by $\hat{X}_n$ and prove that, under the null hypothesis, $\hat{Q}_N \to^d \chi^2_{p^2H}$.

The Monte Carlo simulation of the distribution of $\hat{Q}_N$ and χ^2 distribution, based on 1000 replications for $H=$10,20,30 and $N = 100$, is demonstrated in Figure 1, based on brownian motion processes. It can easily be seen that $\hat{Q}_N$ takes smaller values than those predicted by its asymptotic distribution and the observed distribution of $\hat{Q}_N$ is more skewed than a $\chi^2_{p^2H}$ distribution, for large values of H.

This shortcoming results in introducing two modifications of the Gabrys and Kokoszka statistic, which will be studied in more details in the next section.

3. Modified portmanteau tests

As mentioned in the previous section, in functional time series analysis, the portmanteau test statistic $\hat{Q}_N$ is presented by Gabrys and Kokoszka [13] and, for large values of H, it takes smaller values than those predicted by its asymptotic distribution. Figure 1 suggests that

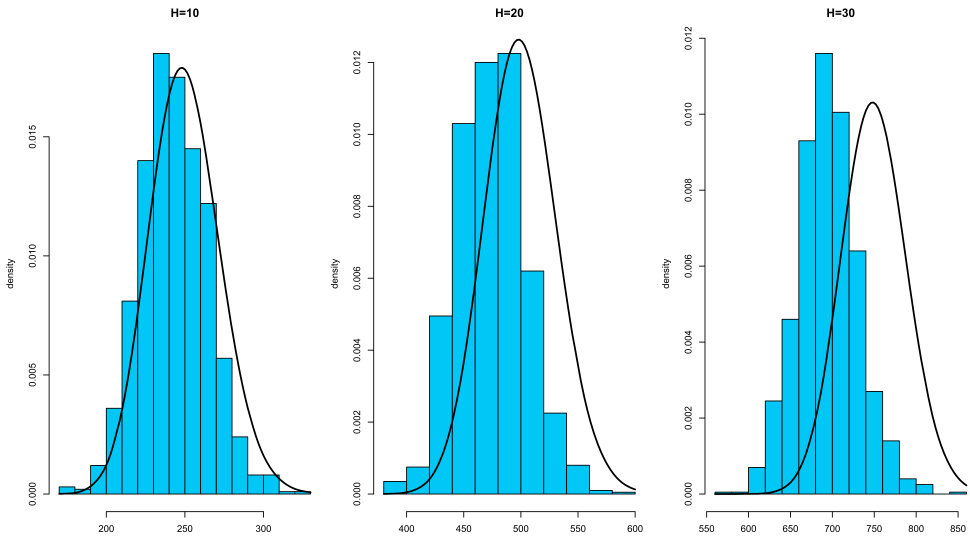

Figure 1. Mont Carlo distribution of $\hat{Q}_N$ and approximation.

moving the empirical distribution of $\hat{Q}_N$ ahead may improve the performance of the test statistic. In this section, we amend the Gabrys and Kokoszka statistic by adding and multiplying a number to present two modifications, which have better small-sample properties. These modifications are based on the characteristics of $\hat{Q}_N$ and the covariance operator of a sequence of iid random functions and resemble the Ljung and Box [2] and Li and McLeod [4] statistics.

The next theorem is devoted to study the expectation and covariance of $c_h(k,l)$, for a sequence of iid random elements.

Theorem 3.1: *Let $\mathbf{X}_t$ be a sequence of iid random functions in $L^2[0,1]$ with mean 0 and covariance operator C with eigenfunctions and eigenvalues ν_j and λ_j, respectively. Define $c_h(k,l)$ as in* (8). *Then,*

$$E\left(c_h\left(k,l\right)\right) = \begin{cases} 0 & h \neq 0 \\ \lambda_k \delta_{kl}, & h = 0 \end{cases}, \tag{10}$$

and, for $r \neq 0$,

$$Cov(c_r(i,j), c_h(k,l)) = \frac{N-r}{N^2} \lambda_i \lambda_j \delta_{ik} \delta_{jl} \delta_{rh}, \tag{11}$$

where δ_{kl} is the kronecker delta.

Proof: Using equation (8), we have

$$\begin{aligned} E\left(c_h(k,l)\right) &= \frac{1}{N} \sum_{t=1}^{N-h} E\left\langle X_t, \nu_k \right\rangle \left\langle X_{t+r}, \nu_l \right\rangle \\ &= \frac{1}{N} \sum_{t=1}^{N-h} \left\langle C_h\left(\nu_k\right), \nu_l \right\rangle . \end{aligned}$$

Since $\mathbf{X_t}$'s are independent, $E(c_h(k,l)) = 0$, for $h \neq 0$, and consequently,

$$E\left(c_h\left(k,l\right)\right) = \frac{N-0}{N}\left\langle C_0\nu_k, \nu_l\right\rangle \delta_{0h} = \lambda_k \delta_{kl}\delta_{0h}.$$

For $r \neq 0$, we have

$$\begin{aligned} Cov(c_r(i,j), c_h(k,l)) &= E(c_r(i,j)\ c_h(k,l)) \\ &= \frac{1}{N^2}\sum_{t=1}^{N-r}\sum_{m=1}^{N-h} E\left\langle X_t, \nu_i\right\rangle \left\langle X_{t+r}, \nu_j\right\rangle \left\langle X_m, \nu_k\right\rangle \left\langle X_{m+h}, \nu_l\right\rangle. \end{aligned}$$

Since X_t's are independent, for $m \neq t$, $Cov(c_r(i,j), c_h(k,l)) = 0$ and, for $m = t$ and $h \neq r$, $Cov(c_r(i,j), c_h(k,l)) = 0$. For $m = t$ and $h = r$, we have

$$Cov\left(c_r\left(i,j\right), c_h\left(k,l\right)\right) = \frac{1}{N^2}\sum_{t=1}^{N-r} E\left\langle X_t, \nu_i\right\rangle \left\langle X_t, \nu_k\right\rangle \left\langle X_{t+r}, \nu_j\right\rangle \left\langle X_{t+r}, \nu_l\right\rangle.$$

Since $r \neq 0$,

$$Cov\left(c_r\left(i,j\right), c_h\left(k,l\right)\right) = \frac{N-r}{N^2}\left\langle C_0(\nu_i), \nu_k\right\rangle \left\langle C_0\left(\nu_j\right), \nu_l\right\rangle = \frac{N-r}{N^2}\lambda_i\lambda_j\delta_{ik}\delta_{jl}\delta_{rh}.$$

■

The following functional versions of the portmanteau tests corroborate the Ljung and Box [2] and Li and McLeod [4] statistics and are denoted by Q_N^{LB} and Q_N^{LM}, respectively:

$$Q_N^{LB} = \sum_{h=1}^{H}\sum_{i,j=1}^{p}\frac{N^2}{N-h}\ r_{f,h}(i,j)\ r_{b,h}(i,j) \tag{12}$$

$$Q_N^{LM} = Q_N + \frac{p^2H(H+1)}{2N}. \tag{13}$$

Note that the multiplier in Q_N^{LB} is defined based on the results of theorem 3.1. The following theorems establish the limiting distribution of the test statistics $\hat{Q}_N^{LB}$ and $\hat{Q}_N^{LM}$, under the null hypothesis.

Theorem 3.2: *Let Assumption 1 holds. Under the null hypothesis, H_0, the test statistic $\hat{Q}_N^{LB}$ converges to some chi-square distribution, i.e.*

$$\hat{Q}_N^{LB} \to^d \chi^2_{p^2H}. \tag{14}$$

Proof: Based on the theorem B.3 of Gabrys and Kokoszka, we have

$$N\sum_{i,j=1}^{p} r_{f,h}(i,j)\ r_{b,h}(i,j) \to^d \chi^2_{p^2}. \tag{15}$$

Since $(N/(N-h)) \to 1$, applying the Slutsky's theorem, it can be concluded $\hat{Q}_N^{LB} \to^d \chi^2_{p^2H}$. ■

Theorem 3.3: *Let Assumption* 1 *holds. Then, under the null hypothesis,* $\hat{Q}_N^{LM} \to^d \chi^2_{p^2H}$.

Proof: By the theorem B.3 of Gasbrys and Kokoszka, $Q_N \to^d \chi^2_{p^2H}$. On the other hand, $(p^2H(H+1)/2N) \to 0$. So, using the Slutsky's theorem, it can be shown $\hat{Q}_N^{LM} \to^d \chi^2_{p^2H}$. ■

4. Portmanteau tests in functional autoregressive processes of order one

Among various functional time series models, functional autoregressive processes are of great importance. These processes, which were studied in details by Bosq [15], have been used in several interesting applications (see, e.g. Laukaitis and Rackauskas [16]; Fernndez de Castro, Guillas, and Gonzles Manteiga 2005; Kargin and Onatski [17]).

Definition 4.1: A sequence $\mathbf{X} = (X_n, n \in \mathbb{Z})$ of zero-mean functional random variables is called an functional autoregressive processes of order 1 (ARH(1)) associated with $(0, \boldsymbol{\varepsilon}, \rho)$ if it is stationary and such that

$$X_{n+1} = \rho X_n + \varepsilon_n, \tag{16}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_n, n \in \mathbb{Z})$ is a functional H-white noise process and ρ is a bounded linear operator on $L^2[0,1]$.

By Theorems 3.4 and 3.6 of Bosq [15], it can easily be shown that $\mathbf{X}_n = (X_{1n}, X_{2n}, \ldots, X_{pn})'$ form a stationary VAR(1) process, i.e.

$$X_{k,n+1} = \sum_{i=1}^{p} \rho_{ki} X_{in} + \varepsilon_{k,n+1}, \quad k = 1, 2, \ldots, p, \tag{17}$$

where $\rho_{ik} = \langle v_i, \rho v_k \rangle$ and $\varepsilon_{k,n} = \langle \epsilon_n, v_k \rangle$, or equivalently, in the vector form, $\mathbf{X}_{n+1} = \boldsymbol{\rho}\mathbf{X}_n + \boldsymbol{\varepsilon}_{n+1}$.

The following theorems establish the consistency of statistics against the functional AR(1) model.

Theorem 4.1: *Let the functional observations X_n follow a stationary solution of a standard functional AR(1) process, Assumption* 1 *holds, and p is chosen so large that the $p \times p$ matrix $\boldsymbol{\rho}$ is not zero. Then $\hat{Q}_N^{LB} \to^P \infty$.*

Proof: By Theorem 2 of Gasbrys and Kokoszka, we have:

$$\hat{Q}_N \to^P \infty, \tag{18}$$

i.e. for any arbitrary large number $k > 0$, $\lim_{N\to\infty} P(\hat{Q}_N > k) = 1$. Therefore,

$$\begin{aligned} P\left(\hat{Q}_N^{LB} > k\right) &\leq P\left(\frac{N}{N-H}\hat{Q}_N > k\right) \\ &= P\left(\hat{Q}_N > \frac{k\,(N-H)}{N}\right) \to 1 \end{aligned}$$

and the proof is completed. ■

The following theorem, which demonstrates the consistency of $\hat{Q}_N^{LM}$ against the functional AR(1) model, can be proved using the same method as in Theorem 4.1.

Theorem 4.2: *Suppose that the functional observations X_n follow a stationary solution of a standard ARH(1) process, Assumption* 1 *holds, and p is so large that the $p \times p$ matrix $\boldsymbol{\rho}$ is not zero. Then $\hat{Q}_N^{LM} \to^P \infty$.*

5. Simulation study and real data analysis

This section is concern with simulation and real data analysis to demonstrate the superiority of the modified portmanteau statistics.

5.1. Simulation study

To determine the empirical size, independent trajectories of the standard Brownian motion (BM) on $[0, 1]$ and the standard Brownian bridge (BB) is generated. This is done by transforming cumulative sums of independent normal variables, computed on a grid of 100 equispaced points in $[0, 1]$. The B-spline functional bases is applied and 49 basis functions is used. All results are based on 1000 replications.

Table 1 shows the Kolmogorov distance, the sup-distance between the empirical distribution functions of test statistics when the observations are BM process and limiting (Chi-square) distribution for several values of the lag $H = 1,3,5,7$; and sample sizes $N = 25, 50, 100, 200$. In the all case, the Kolmogorov distance tends to zero as sample size N increase. For fixed sample size, we see deficiency of Q_N for $H > 1$. This shortcoming is more clear when sample size N is small. While the distance of statistics Q_N^{LB} and Q_N^{LM} seems to be constant with respect to H. We see more closeness between empirical distribution of Q_N^{LB} and corresponding limiting distribution than Q_N and Q_N^{LM}. Figure 2 displays the kernel density plots of test statistics and corresponding limiting distributions and confirms the results of Table 1.

Tables 2 and 3 show the percentage of $Q > \chi^2(1 - \alpha, df)$ for $\alpha = 0.01, 0.05, 0.10$ and $Q = Q_N, Q_N^{LB}, Q_N^{LM}$ for the BM and BB and several values of the lag $H = 1, 3, 5, 7$; and sample sizes $N = 25,50,100,200$. In most cases, the size improves somewhat with the increase of N. We see smaller type I error of Q_N. This is not surprising, since Q_N is smaller in value than Q_N^{LM} and Q_N^{LB}.

Table 1. Kolmogorov distance between the empirical cumulative distribution of the test statistics and the cumulative limit distribution when the observations are brownian motion.

N	Lag	Q_N	Q_N^{LB}	Q_N^{LM}	N	Lag	Q_N	Q_N^{LB}	Q_N^{LM}
	1	0.083	0.096	0.103		1	0.019	0.032	0.030
25	3	0.192	0.102	0.121	100	3	0.054	0.040	0.042
	5	0.366	0.104	0.130		5	0.091	0.048	0.052
	7	0.555	0.089	0.132		7	0.142	0.06	0.072
	1	0.024	0.028	0.034		1	0.034	0.028	0.029
50	3	0.088	0.060	0.070	200	3	0.045	0.025	0.024
	5	0.170	0.072	0.088		5	0.069	0.022	0.027
	7	0.290	0.075	0.085		7	0.107	0.033	0.036

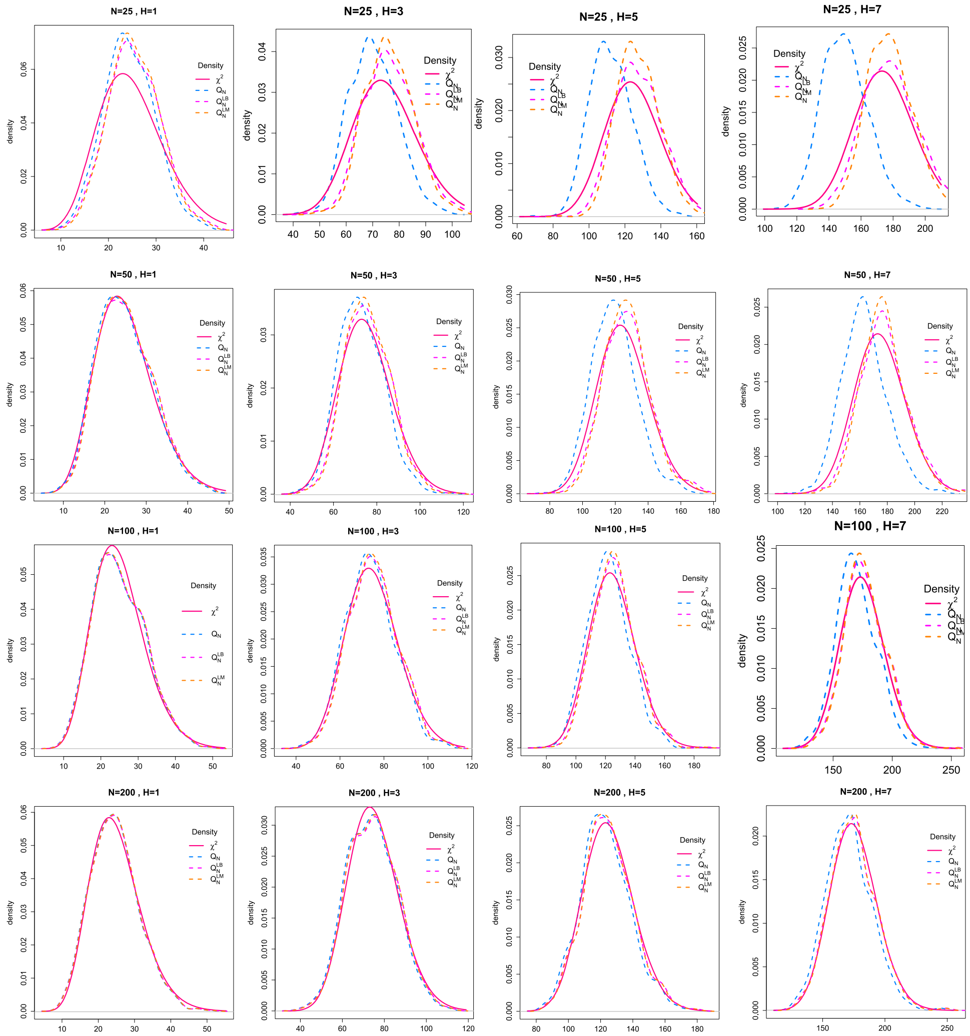

Figure 2. Empirical density(kernel) of statistics (dashed lines) and limiting distribution (solid lines).

In order to compare the power of the test statistics, we focus on the ARH(1) model, which can be written as

$$X_t\left(u\right) = \int_0^1 \psi\left(s, u\right) X_{t-1}\left(s\right) \mathrm{d}s + B_t\left(u\right), \tag{19}$$

where ψ supposed the Gaussian kernel,

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

with the choice of $C = 0.3416$, and B_t is BM innovation. Tables 4 shows the percentage of $Q > \chi^2(1 - \alpha, df)$ for $\alpha = 0.01, 0.05, 0.10$ and $Q = Q_N, Q_N^{LB}, Q_N^{LM}$ for ARH(1) model (19) and several values of the lag $H = 1,3,5,7$; and sample sizes $N = 25,50$. The results for $N = 100,200$ are eliminated since they was practically 100% for and all choices of other

Table 2. Empirical size of the test when the observations are independent Brownian motions.

		Q_N			Q_N^{LB}			Q_N^{LM}		
N	Lag	10%	5%	1%	10%	5%	1%	10%	5%	1%
	1	6.5	2.1	0	8.5	3.2	0	7.6	2.4	0.0
25	3	2.8	1.2	0.0	6.6	2.7	0.1	5.4	2.3	0.1
	5	1.6	0.7	0.0	8.0	3.9	0.7	6.2	2.2	0.4
	7	0.9	0.2	0.0	10.4	5.4	1.2	7.3	3.2	0.3
	1	8.7	4.0	0.5	9.7	4.9	0.5	9.4	4.5	0.5
50	3	5.7	2.3	0.0	8.6	3.6	0.6	7.8	3	0.4
	5	5.5	2.4	0.5	9.7	5.1	0.9	8.6	4.2	0.8
	7	4.3	1.6	0.3	9.9	6.1	1.3	8.9	4.4	0.6
	1	9.5	4.4	0.4	10.2	4.6	0.7	10	4.5	0.5
100	3	9.1	4.3	0.5	10.3	4.9	0.8	10	4.5	0.7
	5	7.1	2.9	0.4	9.1	4.7	0.5	8.8	4.1	0.4
	7	5.3	2.4	0.5	7.5	4.1	0.9	6.9	3.6	0.8
	1	10.8	5	1.1	11.6	5.1	1.2	11.1	5	1.1
150	3	9.4	5.0	1.1	10.0	5.3	1.2	9.8	5.2	1.2
	5	8.3	4.7	1.0	9.8	5.5	1.2	9.7	5.3	1.1
	7	7.2	3.2	0.2	10.2	4.8	0.9	9.4	4.4	0.9
	1	10.9	4.2	1.2	10.9	4.3	1.3	10.9	4.3	1.2
200	3	9.6	4.5	0.5	10.4	5.1	0.5	10.3	5.0	0.5
	5	8.1	3.8	0.6	9.0	4.2	0.8	8.7	4.0	0.7
	7	6.6	3.0	0.5	7.9	4.1	0.6	7.3	4.0	0.6

Table 3. Empirical size of the test when the observations are independent Brownian bridge.

		Q_N			Q_N^{LB}			Q_N^{LM}		
N	Lag	10%	5%	1%	10%	5%	1%	10%	5%	1%
	1	6.6	2.4	0.1	8.9	3.5	0.4	7.7	3	0.1
25	3	4.2	1.6	0.0	8.9	3.6	0.4	8	2.5	0.2
	5	2.1	0.9	0.0	9.8	4.6	0.8	6.9	2.6	0.4
	7	1.0	0.6	0.0	9.7	5.9	1.0	7.4	3.1	0.6
	1	8.0	3.5	0.5	8.7	3.6	0.6	8.3	3.5	0.5
50	3	5.7	2.7	0.3	8.6	3.9	0.8	7.6	3.3	0.5
	5	4.8	2.0	0.2	9.1	4.8	0.7	8.0	3.8	0.7
	7	3.4	1.7	0.5	10.3	4.7	1.1	8.9	3.7	0.6
	1	10.5	4.0	0.7	11.0	4.3	0.7	10.6	4.2	0.7
100	3	9.5	4.4	0.6	10.7	5.5	0.9	10.5	5	0.7
	5	7.6	3.5	0.7	10.2	5.1	1	9.4	4.7	1
	7	6.4	3.3	0.8	10.9	5.1	1.2	10.2	4.7	1.1
	1	10.9	4.9	0.8	10.9	5.2	1	10.9	5.2	0.9
150	3	8.8	3.4	0.9	9.7	4.6	1.0	9.6	4.3	1
	5	8.1	3.7	0.6	10.7	5.2	0.7	10.4	4.9	0.7
	7	7.5	2.9	0.4	9.1	4.9	0.6	8.7	4.4	0.6
	1	10.3	5	0.9	10.5	5.2	0.9	10.4	5.2	0.9
200	3	9.2	4.5	1	9.8	4.7	1.1	9.8	4.6	1
	5	8.3	4	0.6	9.6	4.4	0.6	9.3	4.1	0.6
	7	7.3	3.5	0.3	9.2	4.4	0.3	8.7	4	0.3

Table 4. Empirical power of the test when the observations are ARH(1) with Brownian motion innovations.

		Q_N			Q_N^{LB}			Q_N^{LM}		
N	Lag	10%	5%	1%	10%	5%	1%	10%	5%	1%
	1	0.902	0.814	0.455	0.917	0.849	0.537	0.911	0.825	0.489
25	3	0.721	0.597	0.346	0.793	0.695	0.455	0.784	0.664	0.41
	5	0.512	0.384	0.203	0.678	0.573	0.349	0.655	0.549	0.308
	7	0.341	0.27	0.129	0.606	0.499	0.317	0.596	0.48	0.281
	1	0.902	0.814	0.455	0.917	0.849	0.537	0.911	0.825	0.489
50	3	0.721	0.597	0.346	0.793	0.695	0.455	0.784	0.664	0.41
	5	0.512	0.384	0.203	0.678	0.573	0.349	0.655	0.549	0.308
	7	0.341	0.27	0.129	0.606	0.499	0.317	0.596	0.48	0.281

parameters. As one can see, Q_N^{LM} and Q_N^{LB} have more power than Q_N specially when sample size is small.

5.2. Application to online retail data

The second application is a transnational dataset from Chen et al. (2012) which contains all the transactions occurring between 1 December 2010 and 9 December 2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts and many customers of the company are wholesalers. Many customers of the company are wholesalers. The dataset is available in UCI repository datasets. Here, the variability of volumes of transactions over 5 min intervals from 8:00 to 18:00 in the first 180 days of 2011 is considered. To do this, the data into considered time intervals are aggregated to obtain a data matrix. More precisely, $V_t(u_i)$ is volumes of selling's transactions in the interval u_i of t-th day, $t = 1, \ldots, 180,\ i = 1, \ldots, 180$. To remove the weekly periodicity, we worked with the differences

$$X_t(u_i) = V_{t+7}(u_i) - V_t(u_i),\ t = 1, \ldots, 173. \tag{20}$$

To stabilize the variance, the square root transformation is used. The observed data transformed to functional objects using 39 Spline basis. Figure 3 displays the data.

Figure 4 shows the p-value of the tests using statistics Q_N, Q_N^{LM} and Q_N^{LB}. As we can see, although all tests reject the null hypothesis for small lags, for large lags, Q_N can no longer reject the null hypothesis.

5.3. Application to diurnal geomagnetic variation data

In this section, we apply our test to ground-based magnetogram records taken from INTERMAGNET institutes collected at Honolulu station, in 2015. These data are considered to find out the structure of this important complex geosystem. Following Gabrys and Kokoszka [13], here we focus only on the horizontal intensity measure. 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. This intensity taken every minute, giving 1440 total measurements per day.

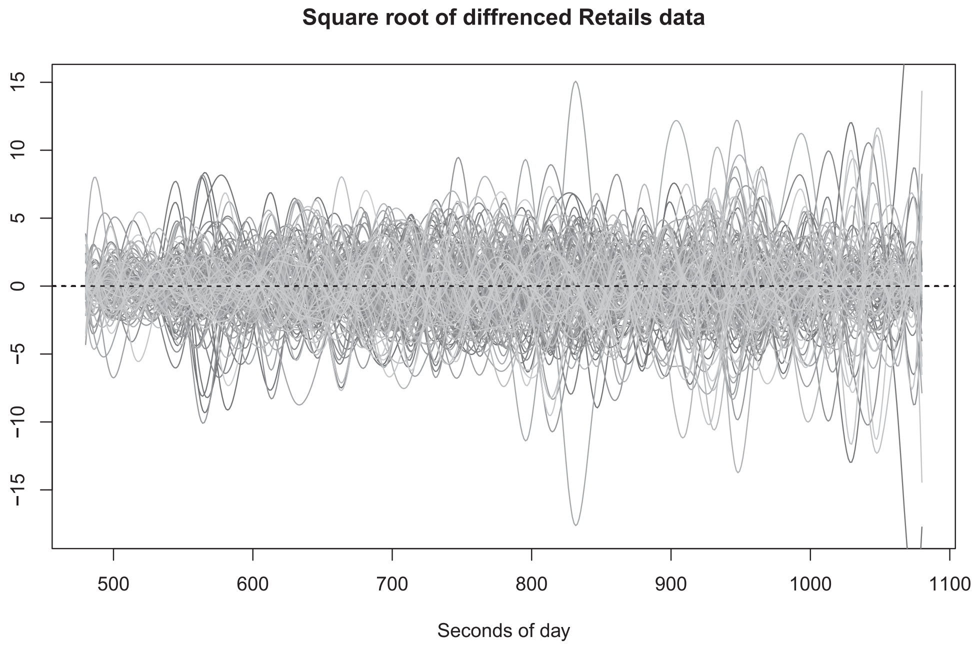

Figure 3. The Online Retail dataset (square root of weekly differences).

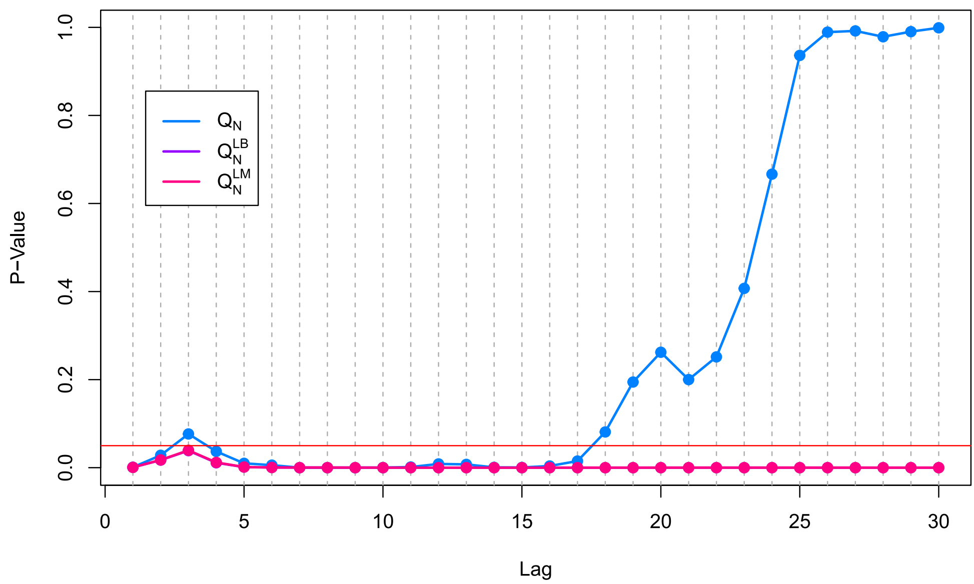

Figure 4. The *p*-value of Portmanteau tests Q_N and Q_N^{LM} for Retail dataset.

We considered each day as a functional observation and 130 of such functional taken from 26 February to 6 July 2015. Figure 5 shows this dataset which is smoothed using 99 B-spline orthogonal basis system. It is argued that the curves follow a dependent dynamic pattern and we expect to see rejection of the null hypothesis by the considered tests. As

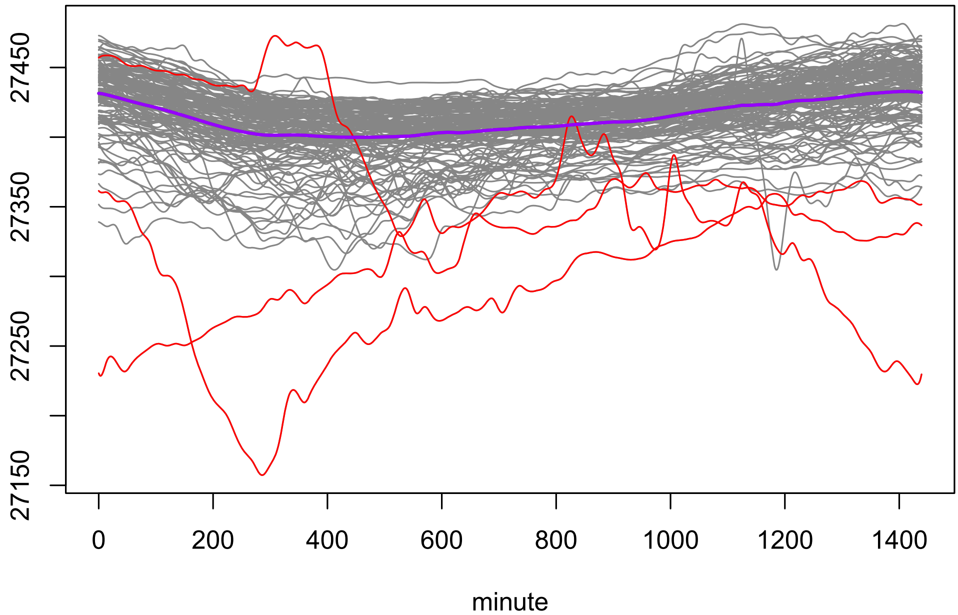

Figure 5. The horizontal components of Honolulu dataset from 26 February 2015 to 6 July 2015.

can be seen in Figure 5, it seems that the dataset contains three extreme functional curves. Considering the null hypothesis and using the R Package fda.usc, the existance of three outliers is confirmed. In order to analyse this dataset and investigate the effect of these outliers, we performed the proposed tests before and after imputing extremes using moving average interpolation.

Figure 6 shows the p-value of the tests using statistics Q_N, Q_N^{LM} and Q_N^{LB}. As can be seen, when considering the actual data set, all tests reject the null hypothesis for small lags and, for large lags (lags greater than 38), Q_N can no longer reject the null hypothesis. However, when the outliers are imputed using moving average interpolation, Q_N^{LM} and Q_N^{LB} have exactly the same performance and Q_N reject the null hypothesis up to lag 63.

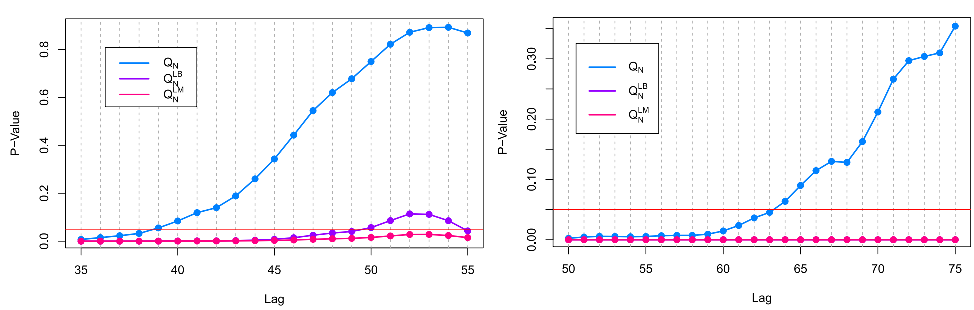

Figure 6. The p-value of Portmanteau tests Q_N and Q_N^{LM}, before (left) and after (right) replacing outliers.

Acknowledgments

The results presented in this paper rely on the data collected at magnetic observatories. The authors thank the National Institutes that support them and INTERMAGNET for promoting high standards of magnetic observatory practice (www.intermagnet.org).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Box GE, Pierce DA. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J Am Stat Assoc. 1970;65(332):1509–1526.
[2] Ljung GM, Box GE. On a measure of lack of fit in time series models. Biometrika. 1978;65(2):297–303.
[3] Hosking JR. The multivariate portmanteau statistic. J Am Stat Assoc. 1980;75(371):602–608.
[4] Li WK, McLeod AI. Distribution of the residual autocorrelations in multivariate ARMA time series models. J R Stat Soc Ser B (Methodological). 1981;43(2):231–239.
[5] McLeod AI, Li WK. Diagnostic checking ARMA time series models using squared-residual autocorrelations. J Time Ser Anal. 1983;4(4):269–273.
[6] Monti AC. A proposal for a residual autocorrelation test in linear models. Biometrika. 1994;81(4):776–780.
[7] Chitturi RV. Distribution of residual autocorrelations in multiple autoregressive schemes. J Am Stat Assoc. 1974;69(348):928–934.
[8] Francq C, Raïssi H. Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors. J Time Ser Anal. 2007;28(3):454–470.
[9] Mahdi E, McLeod AI. Improved multivariate portmanteau test. J Time Ser Anal. 2012;33(2):211–222.
[10] Ramsay JO, Silverman BW. Functional data analysis. New York: Springer; 2005.
[11] Ferraty F, Vieu P. Nonparametric functional data analysis: theory and practice. New York: Springer Science and Business Media; 2006.
[12] Horváth L, Kokoszka P. Inference for functional data with applications (Vol. 200). New York: Springer Science and Business Media; 2012.
[13] Gabrys R, Kokoszka P. Portmanteau test of independence for functional observations. J Am Stat Assoc. 2007;102(480):1338–1348.
[14] Li Y, Wang N, Carroll RJ. Selecting the number of principal components in functional data. J Am Stat Assoc. 2013;108(504):1284–1294.
[15] Bosq D. Linear processes in function spaces: theory and applications. New York: Springer-Verlag; 2000.
[16] Laukaitis A, Rackauskas A. Functional data analysis of payment systems. Nonlinear Anal: Model Control. 2002;7:53–68.
[17] Kargin V, Onatski A. Curve forecasting by functional autoregression. J Multivar Anal. 2008;99:2508–2526.