

## NONPARAMETRIC TESTS FOR SERIAL INDEPENDENCE BASED ON QUADRATIC FORMS

Cees Diks and Valentyn Panchenko

*University of Amsterdam and University of New South Wales*

*Abstract:* Tests for serial independence and goodness-of-fit based on divergence notions between probability distributions, such as the Kullback-Leibler divergence or Hellinger distance, have recently received much interest in time series analysis. The aim of this paper is to introduce tests for serial independence using kernel-based quadratic forms. This separates the problem of consistently estimating the divergence measure from that of consistently estimating the underlying joint densities, the existence of which is no longer required. Exact level tests are obtained by implementing a Monte Carlo procedure using permutations of the original observations. The bandwidth selection problem is addressed by introducing a multiple bandwidth procedure based on a range of different bandwidth values. After numerically establishing that the tests perform well compared to existing nonparametric tests, applications to estimated time series residuals are considered. The approach is illustrated with an application to financial returns data.

*Key words and phrases:* Bandwidth selection, nonparametric tests, serial independence, quadratic forms.

### 1. Introduction

Tests for serial independence are important diagnostic tools for time series modelling. A classical example is the Durbin and Watson (1950, 1951) test for serial correlation, while Li (2004) describes a range of more recent diagnostic checks in time series. Because in many applied fields it has been realised that the time series processes encountered are nonlinear and non-Gaussian, nonparametric measures of dependence are currently receiving much interest. In fact nonparametric tests for independence date back to Hoeffding (1948), followed by the work of, among others, Heer (1991), Rosenblatt and Wahlen (1992) and Ahmad and Li (1997). However, the methods considered there do not directly provide tests for serial independence in time series. In the time series context several different nonparametric approaches have been taken. For instance, Delgado and Mora (2000) and some references therein use the empirical distribution function, while Hong (2000) suggested a frequency domain approach for testing serial independence. The approach we consider in this paper is more closely related to the nonparametric kernel-based BDS test of Brock, Dechert, Scheinkman and LeBaron

(1996). Before going into details, the hypothesis of serial independence in time series contexts is made explicit.

We consider a strictly stationary real-valued time series process  $\{X_t\}$ ,  $t \in \mathbb{Z}$ . We introduce the  $m$ -dimensional lag  $\ell$  delay vectors  $X_t^{m,\ell} = (X_t, X_{t+\ell}, X_{t+2\ell}, \dots, X_{t+(m-1)\ell})'$  and denote the time invariant probability measure of  $X_t^{m,\ell}$  by  $\nu_m$ , suppressing the dependence on  $\ell$ , which we think of as being fixed. The null hypothesis of interest is that  $\{X_t\}$  is serially independent, i.e., its elements are independent and identically distributed (i.i.d.). Under the null hypothesis the  $m$ -dimensional delay vector measure  $\nu_m$  is equal to the product measure  $\nu_1^m = \nu_1 \times \dots \times \nu_1$  ( $m$  terms) of marginal probability measures. In cases where the probability density functions (pdfs) of  $X_t^{m,\ell}$  exist, these are denoted by  $f_m(x)$ , where  $x = (x_1, \dots, x_m)'$ , and the null hypothesis can be expressed as  $f_m(x) = f_1(x_1) \times \dots \times f_1(x_m)$ .

Recently, information theoretic measures of divergence such as the Kullback-Leibler divergence and the Hellinger distance gained much attention in the literature, see Granger and Lin (1994), Hong and White (2005), Granger, Maasoumi and Racine (2004). These measures of divergence, being defined in terms of the joint and marginal pdfs, are usually estimated on the basis of plug-in kernel estimators of the joint and marginal densities. To establish consistency of the test statistics thus obtained, it suffices to take the bandwidth according to the optimal value for kernel density estimation as in Silverman (1986), although it has been recognised that this choice need not be optimal in terms of the power of the tests. Along the same lines, Feuerverger (1993) reaches the conclusion that the consistency of the associated density estimator is not required for the consistency of the quadratic measure of Rosenblatt (1975), given by  $T = \int (\hat{f}_1 - \hat{f}_2)^2$ , where  $\hat{f}_i$  are kernel density estimators. Anderson, Hall and Titterington (1994) indicate that relative oversmoothing is appropriate for this type of statistic in a two-sample test. A problem related to using the consistency of the plug-in estimators is the difficulty of producing efficient kernel estimates of multivariate densities due to the curse of dimensionality.

In this paper a different perspective on the above issues is offered by defining divergence measures between distributions using kernel-based quadratic forms. These divergence measures naturally lead to U- and V-statistic estimators (see e.g., Serfling (1980)), which are closely related to the statistic  $T$ . However, it becomes apparent that the bandwidth plays an entirely different role here than in nonparametric density estimation, where it controls the trade-off between the bias and variance of the density estimators  $\hat{f}_i$ . Starting from quadratic forms, a different divergence measure is associated with each fixed bandwidth value. Each of the members of this family of divergence measures, parameterised by the bandwidth, can be estimated consistently using U- or V-statistics based on

a kernel function with that same particular bandwidth, even if the distributions are not continuous (and hence no underlying densities exist). Rather than acting as a bias-variance trade-off parameter, the bandwidth merely controls the length scale in the sample space at which two probability measures are compared.

It follows from the results by Denker and Keller (1983) that under strict stationarity and absolute regularity, the proposed tests for serial independence are consistent against all fixed alternatives. In contrast to some other tests based on quadratic functionals, such as that of Székely and Rizzo (2005), there is no need to impose any conditions on the moments of the time series. Although consistency does not require the bandwidth to vanish with the sample size, the approach still faces the common bandwidth selection problem. This is addressed by implementing a multiple bandwidth procedure along the lines of the approach of Horowitz and Spokoiny (2001).

In Section 2 we introduce the notion of the squared distance measures between probability distributions in terms of kernel-based quadratic forms and derive some of their properties. Section 3 describes a single bandwidth permutation test for serial independence based on squared distances and investigates how the power depends on different values of the bandwidth. Subsequently the multiple bandwidth procedure to deal with the bandwidth selection problem is described. In Section 4 the finite-sample performance against fixed and local alternatives is compared with some other nonparametric tests for serial independence. There we also investigate the behaviour of the test when applied to estimated residuals. After an application to the log-return series of S&P 500 stock index in Section 5, Section 6 summarises and concludes.

## 2. Quadratic Forms and Their Estimators

In this section we briefly review the distance notions between probability measures in  $\mathbb{R}^m$  that will serve as the divergence between the joint probability measure  $\nu_m$  and its counterpart  $\nu_1^m$ . These functionals were first introduced by Diks, van Zwet, Takens and DeGoede (1996) in the context of measuring the divergence between chaotic time series, and later applied in a test for symmetry by Diks and Tong (1999). For two  $m$ -dimensional probability measures  $\mu_1$  and  $\mu_2$ , consider a quadratic form of the type:

$$Q = \|\mu_1 - \mu_2\|^2 = (\mu_1 - \mu_2, \mu_1 - \mu_2) = (\mu_1, \mu_1) - 2(\mu_1, \mu_2) + (\mu_2, \mu_2),$$

where  $(\mu_i, \mu_j) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_h(x - y) \mu_i(dx) \mu_j(dy)$  is a bilinear form, which can be concisely written as  $(\mu_i, \mu_j) = E(K_h(X - Y))$  where  $X$  and  $Y$  are two independent  $m$ -dimensional vectors, distributed according to  $\mu_i$  and  $\mu_j$ , respectively. Whenever  $K_h(\cdot)$  is a positive definite kernel function this bilinear form defines an inner product, and the squared distance  $Q$  defines a metric on the space

of probability measures on  $\mathbb{R}^m$ . We typically consider kernels that factorise as  $K_h(x) = \prod_{i=1}^m \kappa(x_i/h)$  where  $x_i$  refers to the  $i$ 'th element of vector  $x$ ,  $\kappa(\cdot)$  is a one-dimensional kernel function, which is symmetric around zero, and  $h$  is a bandwidth parameter.

Because Fourier transforms leave the  $L_2$  norm invariant by Parseval's identity, and convolution amounts to multiplication in Fourier space, the quadratic form can be expressed as

$$\begin{aligned} Q &= \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_h(x-y)(\mu_1 - \mu_2)(dx)(\mu_1 - \mu_2)(dy) \\ &= \int_{\mathbb{R}^m} \tilde{K}_h(k)|(\tilde{\mu}_1 - \tilde{\mu}_2)(k)|^2 dk, \end{aligned}$$

where  $\tilde{K}_h(\cdot)$  is the Fourier transform of  $K_h(\cdot)$ ,  $\tilde{\mu}_i$  the characteristic function of  $\mu_i$ , and  $|\cdot|$  the modulus. It follows that if  $\tilde{K}_h(\cdot)$  is an integrable real-valued positive function,  $K_h(\cdot)$  is positive definite, and  $Q = 0$  if and only if  $\mu_1$  and  $\mu_2$  are identical probability measures, and is strictly positive otherwise. Here we focus on three specific cases of positive definite kernels: the Gaussian kernel  $\kappa(x) = \exp(-x^2/4)$ , as in Diks and Tong (1999), the double exponential kernel  $\kappa(x) = \exp(-|x|/4)$ , and the Cauchy kernel  $\kappa(x) = 1/(1+x^2)$ . The factor 4 in the Gaussian and double exponential kernels is chosen for convenience as it simplifies some of the derivations discussed below.

The squared distance  $Q$  satisfies all the essential "ideal" properties of a dependence measure summarised by Granger, Maasoumi and Racine (2004). It is well-defined for continuous as well as discrete random variables. It is nonnegative, equal to zero only in the case of independence, and can be related to the correlation coefficient  $\rho$  in the case of a bivariate normal distribution, as shown in Appendix A (Appendices are provided in an on-line supplement at <http://www3.stat.sinica.edu.tw/statistica>). Since  $(\cdot, \cdot)$  defines an inner product on the space of measures on  $\mathbb{R}^m$ ,  $Q^{1/2}$  is a real distance notion between probability measures with the usual properties, such as the triangular inequality. Although  $Q$  is not invariant under monotonic transformations of marginals, if desired, invariance of estimators can always be achieved by transforming the data to a known marginal distribution, e.g., by using empirical probability integral transforms. Moreover, in Appendix B we establish the equivalence of the quadratic form  $Q$  and the quadratic measure of Rosenblatt (1975).

For convenience we introduce the short-hand notation  $Q^{ij} = (\mu_i, \mu_j)$ . As shown above,  $Q^{ij}$  can be expressed in terms of averages of the kernel function:  $Q^{ij} = E(K_h(X-Y))$  for independent vectors  $X \sim \mu_i$  and  $Y \sim \mu_j$ . This suggests estimating  $Q^{ij}$  using empirical averages of the values of the kernel function obtained from the data, thus leading naturally to the use of U- and V-statistics as

discussed in detail by Serfling (1980). For example, given an observed time series  $\{X_t\}_{t=1}^T$ , from which  $n = T - (m - 1)\ell$  delay vectors  $X_t^{m,\ell}$ ,  $t = 1, \dots, n$  of dimension  $m$  can be constructed, the first term  $Q^{11}$  is estimated by the U-statistic estimator

$$\begin{aligned}\widehat{Q}^{11} &= \frac{2}{n(n-1)} \sum_{t=2}^n \sum_{s=1}^{t-1} K_h(X_t^{m,\ell} - X_s^{m,\ell}) \\ &= \frac{2}{n(n-1)} \sum_{t=2}^n \sum_{s=1}^{t-1} \prod_{k=0}^{m-1} \kappa \left( \frac{X_{t+k\ell} - X_{s+k\ell}}{h} \right).\end{aligned}$$

For the bounded kernel functions considered here, it follows from the work of Denker and Keller (1983), Theorem 1, part (c), that under strict stationarity and absolute regularity of the time series, both U- and V-statistics are consistent and asymptotically normal. In particular this implies  $\widehat{Q}^{11} \xrightarrow{p} Q^{11}$ . Similarly one can construct a consistent U-statistic estimator  $\widehat{C}_h(x) = \frac{1}{n} \sum_{t=1}^n \kappa((x - X_t)/h)$  for  $E[\kappa((x - X)/h)]$  and use this to obtain consistent estimators for  $Q^{12}$  and  $Q^{22}$  after writing these in terms of  $E[\kappa((x - X)/h)]$ :

$$\begin{aligned}\widehat{Q}^{12} &= \frac{1}{n} \sum_{t=1}^n \prod_{k=0}^{m-1} \widehat{C}_h(X_{t+k\ell}), \\ \widehat{Q}^{22} &= \frac{1}{n^m} \prod_{k=0}^{m-1} \left( \sum_{t=1}^n \widehat{C}_h(X_{t+k\ell}) \right).\end{aligned}$$

Taken together,  $\widehat{Q} = \widehat{Q}^{11} - 2\widehat{Q}^{12} + \widehat{Q}^{22}$  is a consistent estimator of  $Q$ .

Note that there is a connection with the BDS test for serial independence by Brock, Dechert, Scheinkman and LeBaron (1996). Using the functional  $Q^{11} - Q^{22}$  with kernel function  $\kappa(x) = I_{(-h,h)}(x)$ , which is 1 if  $x \in (-h, h)$  and 0 otherwise, will lead to the BDS test, with  $Q^{11}$  playing the role of the correlation integral and  $Q^{22}$  of its value under the null hypothesis of serial independence.

Based on the theory of U-statistics one might develop asymptotic theory for the functional  $Q$ , possibly with a suitably chosen rate at which  $h$  tends to zero as  $n \rightarrow \infty$  (cf., Wolff (1994)). However, as reported by Skaug and Tjøstheim (1993), Granger, Maasoumi and Racine (2004) and Hong and White (2005) in similar testing contexts, asymptotic theory provides rather poor finite sample approximations to the null distributions of the test statistics, and inference based on such tests becomes unreliable. To avoid this problem we proceed with a permutation procedure.

### 3. Permutation Test

The idea to use a permutation test in the context of serial independence dates back to Pitman (1937). Due to the decreasing cost of computing power,

permutation tests have gained increasing attention (for a practical exposition see Good (2000)). Under the condition of exchangeability of the observations a permutation test is exact for any sample size  $n$ , i.e., the rejection rate under the null hypothesis is equal to the nominal size  $\alpha$ . Moreover, Hoeffding (1952) shows that under general conditions permutation tests are asymptotically as powerful as certain related parametric tests.

### 3.1. Single bandwidth

First we consider a standard procedure using a single fixed bandwidth  $h$ . Since deviations from the null lead to positive values of  $Q$ , a test based on this squared distance would reject whenever the estimate  $\widehat{Q}$  is too large. Thus, a one-sided test is appropriate in this context. Conditional on the observed values of the data under the null hypothesis of serial independence, each permutation of the observed data is equally likely. We denote the estimate  $\widehat{Q}$  based on the original data as  $\widehat{Q}_0$ . Under the null the values of  $\widehat{Q}_i$ ,  $i = 0, \dots, B$ , computed using the original data and  $B$  permutations, respectively, are exchangeable. An exact p-value (in that it is uniformly distributed on  $1/(B+1), \dots, 1$  under the null) is calculated as

$$\widehat{p} = \frac{\sum_{i=0}^B I(\widehat{Q}_i > \widehat{Q}_0) + L}{B+1}, \quad (1)$$

where  $I(\cdot)$  denotes the indicator function taking the value 1 if the condition in brackets is true and 0 otherwise. Let  $Z = \sum_{i=0}^B I(\widehat{Q}_i = \widehat{Q}_0) \geq 1$  denote the number of ties plus one. In case  $Z = 1$ ,  $L = 1$ , while for  $Z > 1$ , take  $L$  uniformly distributed on  $1, \dots, Z$ . That is, each rank of  $\widehat{Q}_0$  among the  $\widehat{Q}_i$  that happen to be equal to  $\widehat{Q}_0$ , is taken to be equally probable. This is equivalent to adding a very small amount of noise to each of the  $\widehat{Q}_i$ 's before determining their ranks, thus making the rank of  $\widehat{Q}_0$  among the  $\widehat{Q}_i$  unique. If  $0 < \alpha = k/(B+1) < 1$  for some integer  $k$ , rejecting whenever  $\widehat{p} \leq \alpha$  yields an exact level- $\alpha$  test. Generally, the power of a permutation test decreases if the number of permutations  $B$  decreases. The results by Marriott (1979) indicate that little power is lost by taking  $B+1 = 5/\alpha$ .

Notice that the term  $\widehat{Q}^{22}$  is constant under permutations, and hence can be left out of consideration while determining the significance of  $\widehat{Q}$ . This reflects the fact that  $Q^{22}$  is a functional of the marginal distribution, which plays a role here as an infinite dimensional nuisance parameter.

So far we have only considered the calculation of p-values for a fixed bandwidth. To deal with the problem of bandwidth selection, Subsection 3.3 describes a method for determining a single p-value over a range of bandwidths. How-

ever, we first motivate the multiple bandwidth procedure by presenting some bandwidth-related simulation results.

### 3.2. Bandwidth-related simulations

Hereafter we refer to the bandwidth that yields the highest empirical power for a fixed size  $\alpha$  as the optimal bandwidth  $h^*$ . We investigate the dependence of the optimal bandwidth on three parameters, namely the data generating process (DGP), the delay vector dimension  $m$ , and the sample size  $n$ . A description of the DGPs used, along with broader simulation results, are presented in Section 4. Here we only display bandwidth-related simulations. We consider  $d = 30$  different bandwidth values  $h_i$  ranging from 0.01 to 3.0, equidistant on a logarithmic scale:

$$h_i = h_{\max} \left( \frac{h_{\min}}{h_{\max}} \right)^{\frac{d-i}{d-1}}, \quad i = 1, \dots, d. \quad (2)$$

The number of permutations was set to  $B + 1 = 100$ , including the original series, and the number of simulations was set to 1,000. Since the Cauchy and double exponential kernels gave similar results, we only discuss the results for the Gaussian kernel.

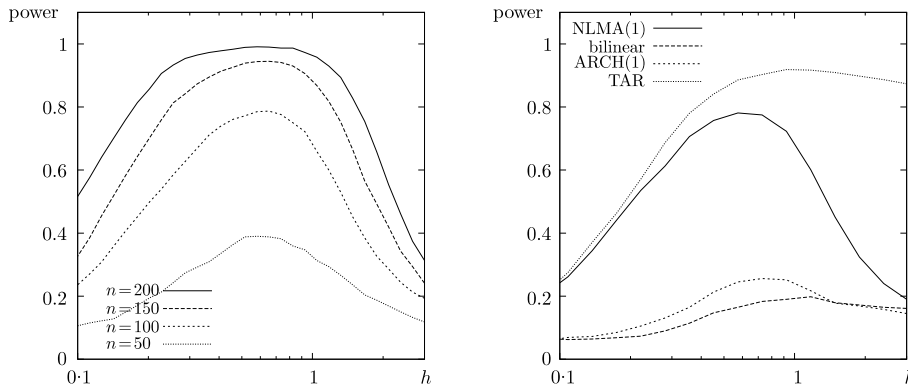


Figure 1. Observed power as a function of bandwidth  $h$ . The left panel shows results, for various series lengths  $n$ , for a nonlinear MA(1) process (NLMA(1), DGP 1); the right panel for various DGPs,  $n = 100$ . In all cases dimension  $m = 2$ , lag  $\ell = 1$ , nominal size  $\alpha = 0.05$ , number of permutations  $B + 1 = 100$ , and number of simulations 1,000.

Figure 1 shows the power as a function of the bandwidth for various lengths  $n$  (left panel, DGP 1,  $m = 2$ ,  $\ell = 1$ ), and for various DGPs (right panel,  $n = 100$ ,  $m = 2$ ,  $\ell = 1$ ). The left panel shows no clear shift in the optimal bandwidth  $h^*$  as  $n$  increases. Similar results were observed for other DGPs. Intuitively, the

reason is that the optimal bandwidth depends on the typical length scale of the differences between the joint delay vector measure  $\nu_m$  and the product measure  $\nu_1^m$ . As long as this length scale is not taken to decrease with  $n$ , the optimal bandwidth may tend to some finite positive value. Analytical support for a fixed optimal bandwidth was reported by Anderson, Hall and Titterton (1994) in a two-sample test based on a statistic of the type  $T = \int (\hat{f}_1 - \hat{f}_2)^2$ . However, the right panel of Figure 1 illustrates that the optimal bandwidth  $h^*$  depends on the particular DGP, e.g., for the nonlinear MA(1) process (DGP 1),  $h^* \simeq 0.7$ , and for the bilinear process (DGP 7),  $h^* \simeq 1.2$ . This suggests that using a single bandwidth value in a practical situation, when the underlying DGP is not known, may not be optimal.

### 3.3. Multiple bandwidth procedure

Motivated by the above findings, we require a procedure that produces a single test statistic (p-value) incorporating a range of bandwidth values. Horowitz and Spokoiny (2001) suggest an adaptive rate-optimal test that uses many different bandwidths. Since the theoretical distribution of their test statistic under the null is not known, they find critical values by simulation. We develop a similar procedure in the Monte Carlo context and implement it in the form of a multiple bandwidth permutation test. The procedure is based on determining the significance of the smallest single-bandwidth p-value over a range of different bandwidths, and can be summarised as follows.

1. Calculate the vector of  $\hat{Q}_{h,0}$ -values for a range of bandwidths:  $h \in H = \{h_1, \dots, h_d\}$ . We define  $h$  on a geometric grid as at (2).
2. Randomly permute the data and calculate a bootstrap vector  $\hat{Q}_{h,1}$ . Repeat this  $B$  times to obtain  $\hat{Q}_{h,i}$  for  $h \in H$  and  $i = 1, \dots, B$ .
3. Transform  $\hat{Q}_{h,i}$  into a p-value:  $\hat{p}_{h,i} = (\sum_{j=0}^B I(\hat{Q}_{h,j} > \hat{Q}_{h,i}) + L)/(B+1)$ , with  $L$  defined as at (1).
4. Select the smallest p-value among all bandwidths and call it  $\hat{T}_i$ :  $\hat{T}_i = \inf_{h \in H} \hat{p}_{h,i}$ .
5. Calculate an overall p-value on the basis of the rank of  $\hat{T}_0$  among the  $\hat{T}_i$ , i.e.,  $\hat{p} = (\sum_{i=0}^B I(\hat{T}_i < \hat{T}_0) + L)/(B+1)$ , using a ties randomisation procedure as at (1).

In Step 3 we pretend each of the permuted series is the originally observed series and determine the corresponding p-values  $\hat{p}_{h,i}$  that would have been obtained for series  $i$  for each of the different bandwidths. In Step 4, for each series the smallest p-value over the different bandwidths is selected (denoted by  $\hat{T}_i$ ,  $i = 0, \dots, B$ ). We finally use the exchangeability of the  $B$  series under the null to calculate an overall p-value by establishing the significance of  $\hat{T}_0$  for the observed data (Step 5). As in the single bandwidth case, the multiple bandwidth procedure yields an



exact  $\alpha$ -level test if the null hypothesis is rejected whenever  $\hat{p} \leq \alpha$ . The power of this multiple-bandwidth procedure depends on the width of the region  $R = [h_{\min}, h_{\max}]$ , the number  $d$  of elements in the bandwidth set  $H$ , and the number of permutations  $B$ . The region  $R$  should be wide enough to contain  $h^*$  for various DGPs. The number of bandwidths  $d$  chosen in  $R$  is important for the power. Taking  $d$  too small risks losing the optimal bandwidth  $h^*$  through the grid. Our simulations suggest that the empirical power of the multiple bandwidth procedure reduces as the bandwidth region  $R$  becomes wider. Therefore in practice we suggest taking  $R = [0.5, 2.0]$ , which includes  $h^*$  for all considered DGPs. For this region reasonable power is achieved using  $d = 5$  bandwidths.

Also the number of permutations  $B + 1$  has an important impact on the power of our multiple bandwidth procedure. We found empirically that the power for this procedure is more sensitive to the number of permutations  $B$  than for the single bandwidth procedure. The reason for this is that the  $\hat{T}_i$  are discrete multiples of  $1/(B + 1)$ , which for small  $B$  leads to many identical  $\hat{T}_i$ -values (ties) and reduces the power. We find that for the considered region  $R = [0.5, 2.0]$  with  $d = 5$ , taking  $B + 1 = 100$  produces good results. These are the parameter values we recommend in practical applications of the test.

#### 4. Test Performance

We next investigate the power of the proposed test, hereafter  $Q$ -test, and compare it with that of similar nonparametric tests such as the BDS test and the recent test of Granger, Maasoumi and Racine (2004), which we refer to as the GMR test. Permutation tests differ from asymptotic tests (based on the derived asymptotic distribution of test statistic) in that the critical value in the former is a random variable. This fact makes the analytic evaluation of its power function difficult. However, Hoeffding (1952) has shown that under certain conditions the random critical value of the permutation test converges in probability to a constant if the number of permutations  $B$  tends to infinity as  $n \rightarrow \infty$ . Relying on this fact, Hoeffding (1952) investigated the large-sample power properties of permutation tests based on a relatively simple test statistic, and demonstrated that under general conditions the permutation tests are asymptotically as powerful as the corresponding parametric tests. In the present context the test statistic is much more complex and therefore we rely heavily on simulations.

##### 4.1. Fixed alternatives

We compare the rejection rates of the tests against fixed alternatives for the following stationary DGPs, where  $\{\varepsilon_t\}$  is an i.i.d. sequence of  $N(0, 1)$  random

variables:

$$\text{DGP0. } Y_t = \varepsilon_t$$

$$\text{DGP1. } Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$$

$$\text{DGP2. } Y_t = \varepsilon_t + 0.6\varepsilon_{t-1}^2 + 0.6\varepsilon_{t-2}^2$$

$$\text{DGP3. } Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}\varepsilon_{t-2}$$

$$\text{DGP4. } Y_t = 0.3Y_{t-1} + \varepsilon_t$$

$$\text{DGP5. } Y_t = 0.8|Y_{t-1}|^{0.5} + \varepsilon_t$$

$$\text{DGP6. } Y_t = \text{sign}(Y_{t-1}) + \varepsilon_t$$

$$\text{DGP7. } Y_t = 0.6\varepsilon_{t-1}Y_{t-2} + \varepsilon_t$$

$$\text{DGP8. } Y_t = \sqrt{h_t}\varepsilon_t, \quad h_t = 1 + 0.4Y_{t-1}^2$$

$$\text{DGP9. } Y_t = \sqrt{h_t}\varepsilon_t, \quad h_t = 0.01 + 0.80h_{t-1} + 0.15Y_{t-1}^2$$

$$\text{DGP10. } Y_t = \begin{cases} -0.5Y_{t-1} + \varepsilon_t, & Y_{t-1} < 1 \\ 0.4Y_{t-1} + \varepsilon_t, & \text{else} \end{cases}$$

$$\text{DGP11. } Y_t = 4Y_{t-1}(1 - Y_{t-1}) \quad 0 < Y_t < 1$$

$$\text{DGP12. } Y_t = 1 + 0.3Y_{t-2} - 1.4Y_{t-1}^2$$

$$\text{DGP13. } Y_t = Z_t + \sigma\varepsilon_t, \quad Z_t = 1 + 0.3Z_{t-2} - 1.4Z_{t-1}^2.$$

The above DGPs, or slight modifications of these, were previously considered by Granger, Maasoumi and Racine (2004), Granger and Lin (1994), Hong and White (2005), Brock, Dechert, Scheinkman and LeBaron (1996) and others. DGP 0 satisfies the null hypothesis and is included to assess the empirical size of the tests. DGPs 1 – 3 are nonlinear MA processes of order 1, 2 and 2 respectively. Granger, Maasoumi and Racine (2004) suggested that a good measure of dependence should reflect the theoretical properties of these MA processes, i.e., zero dependence at lags beyond their nominal values. DGP 4 is a linear AR(1) process, while DGPs 5 and 6 are nonlinear AR(1) processes. The properties of DGP 6 were investigated by Granger and Teräsvirta (1999). DGP 7 is a bilinear process introduced by Granger and Andersen (1978). DGPs 8 and 9 are instances of ARCH(1) and GARCH(1, 1) processes proposed by Engle (1982) and Bollerslev (1986), respectively. The coefficients of the GARCH(1, 1) process are taken close to the corresponding estimates of Bollerslev (1986). DGP 10 is a TAR process proposed by Tong (1978). DGPs 11 and 12 are the logistic map and the Hénon map respectively, generating deterministic chaotic time series, while DGP 13 is the Hénon map with additive Gaussian observational noise  $\sigma\varepsilon_t$ , where  $\sigma$  equals 20 percent of the standard deviation of the clean Hénon process. We used series of length  $n = 100$  (except  $n = 50$  for DGP 6 and  $n = 20$  for

DGPs 11 – 13), and the total number of permutations, including the original series, was set to  $B + 1 = 100$ . The bandwidth set  $H$  included  $d = 5$  different values in the range  $R = [0.5, 2.0]$  after normalising the series to unit variance. The three different kernels mentioned earlier were used for comparison: the Gaussian, double exponential and Cauchy kernels. We considered lags  $\ell = 1, 2$  and 3 for delay vector dimension  $m = 2$ , and extended the delay vector dimension to  $m = 3, 4, 5$  and 10 for lag  $\ell = 1$ . All tests were conducted at a nominal size of  $\alpha = 0.05$ , and the number of simulations was set to 1,000.

Generally, the BDS test statistic is not necessarily positive under the alternative. This was confirmed by simulations for certain alternatives, e.g., for the logistic map (DGP 11) the rejection rate was smaller than the nominal size while using a one-sided test. Therefore, we implemented it as a two-sided test. To make the BDS test comparable with the  $Q$ -test, we applied a similar multiple bandwidth permutation procedure and doubled the number of permutations to  $B + 1 = 200$  to take into account the two-sidedness. The bandwidth range  $R = [0.5, 2.0]$ , which is typical for the BDS test, coincides exactly with that used in the  $Q$ -test. We set the number of bandwidths to  $d = 5$  also for the BDS test.

We used the original routine for the GMR test to compute rejection rates for the considered DGPs. Since their test embeds the likelihood cross validation of Silverman (1986, Sec. 3.4.4), to select optimal bandwidths (determining separate optimal bandwidth values under the null and the alternative), no bandwidth selection was required. For dimensions higher than two we used their “portmanteau” version of the test.

Figure 2 reports the observed rejection rates (at size  $\alpha = 0.05$ ,  $\ell = 1$ ,  $m = 3$ ) for the considered processes for the introduced  $Q$ -test based on the Gaussian kernel, the BDS test and the GMR test. See Appendix C for the numerical values and extended results (higher lags  $\ell$  and dimensions  $m$ ) of these tests and the  $Q$ -tests based on other kernels. As expected, the actual size of all tests is close to the nominal size. The  $Q$ -test yields powers comparable to those obtained using the BDS and GMR procedures and in 8 out of 13 cases outperforms them, i.e., for the nonlinear MA(1) – MA(2), linear, fractional and sign function AR(1) and TAR processes, and the Hénon map without and with the observational noise (DGPs 1, 2, 4 – 6, 10, 12, 13). In absolute terms the power of the  $Q$ -test is smaller for the nonlinear MA(2) and bilinear processes and the logistic map (DGPs 3, 7, 11), but still comparable to that obtained by the best performing test (for a particular DGP). In comparison with the BDS test, the  $Q$ -test shows less power for the ARCH(1) and GARCH(1, 1) processes (DGPs 8 and 9). The GMR test behaves similar to the  $Q$ -test in this situation. Comparing the performance of the  $Q$ -test based on the Gaussian, double exponential and Cauchy kernels, we do not observe large differences (see Appendix C). Therefore, we proceed with the analysis based on the Gaussian kernel only.

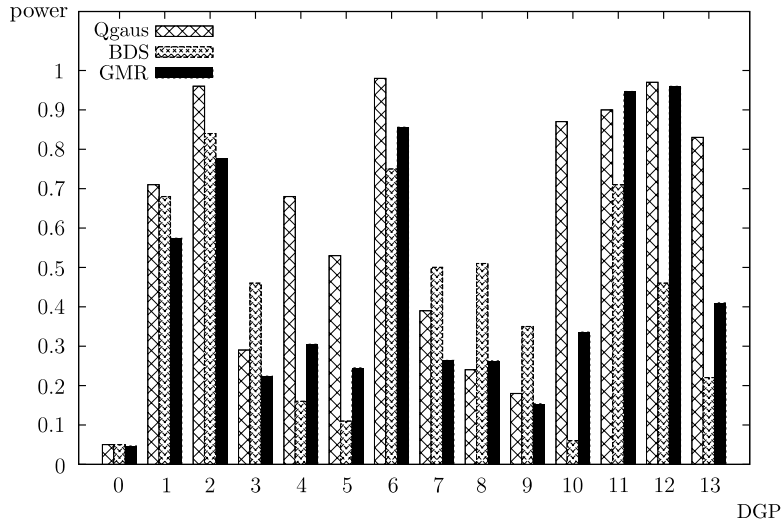


Figure 2. Observed rejection rates (size/power) for various DGPs. Nominal size  $\alpha = 0.05$ , sample size  $n = 100$ , lag  $\ell = 1$ , dimension  $m = 3$ , number of permutations  $B + 1 = 100$  (200 for BDS), number of simulations 1,000.

Table 1. Observed power against GARCH(1, 1) (DGP 9) after (abs) and before (orig) transforming the data to absolute values, nominal size  $\alpha = 0.05$ , sample size  $n = 100$ , lag  $\ell = 1$ , number of permutations  $B + 1 = 100$  (200 for BDS) and number of simulations 1,000.

DGP	Qgaus		BDS		GMR	
	abs	orig	abs	orig	abs	orig
2	0.29	0.13	0.25	0.26	0.12	0.13
3	0.39	0.18	0.31	0.35	0.15	0.15
5	0.46	0.26	0.40	0.43	0.18	0.18
10	0.53	0.38	0.42	0.48	0.15	0.17

The ARCH(1) process (DGP 8) and its generalisation, the GARCH(1, 1) process (DGP 9), are used in financial econometrics to model periods of consecutive large deviations from the mean, interchanged by periods of moderate deviations, mimicking observed behaviour of stock returns. Since the GARCH(1, 1) process is of special interest in financial econometrics we undertake a more detailed analysis of this process. The power of the  $Q$ -test increases if we consider higher delay vector dimensions  $m$  for this DGP. To obtain an even further increase in power against the GARCH(1, 1) process, we can adopt a semi-parametric approach and transform the data to their absolute values before testing. Table 1 shows the rejection rates obtained with the test for GARCH(1, 1) using this transformation in contrast to no transformation. After this transformation the  $Q$ -test becomes

more powerful than the BDS and the GMR test conducted on the transformed and original data. The intuition behind this increase in power lies in the local nature of the kernel. Initially distant delay vectors with differently signed elements can be mapped locally close to each other upon replacing the vector elements by their absolute values, enabling the test to capture more of the dependence. We conclude from this that applying the  $Q$ -test to the absolute values of the data is preferable when structure in volatility is to be detected.

#### 4.2. Local alternatives

We next consider power against local alternatives. For a test similar to that of GMR, Hong and White (2005) found nontrivial power as the distance between the null distribution and a local alternative reduces at the rate  $n^{-1/2}h^{-1/2}$ , with  $h \rightarrow 0$  as required for consistent kernel estimation of the density. The test statistic for the  $Q$ -test is estimated using U-statistics, which in the non-degenerate case converge at the parametric rate  $n^{-1/2}$ . Moreover, the consistency of the  $Q$ -test does not require the bandwidth to diminish with the sample size. Therefore, we expect the test to have nontrivial asymptotic power at the rate  $n^{-1/2}$ , and illustrate this via simulations. For the same reasons a similar rate is expected for the BDS test. Following Hong and White (2005), we consider a sequence of processes with lag  $j$  dependence with the following joint probability function:

$$f_{jn}(y_t, y_{t+j}) = f(y_t)f(y_{t+j})[1 - a_n q_j(y_t, y_{t+j}) + r_{jn}(y_t, y_{t+j})], \quad (3)$$

where  $q_j(y_t, y_{t+j})$  is a function characterising the deviation from the null hypothesis,  $a_n$  governs the rate of convergence to the null as  $n \rightarrow \infty$ , and  $r_{jn}(y_t, y_{t+j})$  is a higher order term obtained from the Taylor series expansion of  $f_{jn}(y_t, y_{t+j})$  around the point  $a_n = 0$ . See Hong and White (2005) for assumptions on  $q_j(\cdot, \cdot)$  and  $r_{jn}(\cdot, \cdot)$  that ensure that  $f_{jn}(\cdot, \cdot)$  is a proper density function.

The simulations are based on an MA(1) process  $Y_t = \varepsilon_t + a_n \varepsilon_{t-1}$  where  $\{\varepsilon_t\}$  is a sequence of i.i.d. standard normal random variables. The joint density of  $(Y_t, Y_{t+1})$  can be represented in the form (3) with  $q_j(y_t, y_{t+j}) = y_t y_{t+j}$ . Figure 3 (left panel) shows the rejection rates (powers) of the considered test against a sequence of local alternatives that converges to the null at the usual parametric rate  $a_n = Cn^{-1/2}$ , where  $C$  is a constant and  $n = 100, \dots, 5,000$ . A horizontal line in the graph would indicate the parametric rate. After an initial transient period for small  $n$ , the curves level out, suggesting that all tests asymptotically approach the parametric rate. The  $Q$ -test has a substantially larger nontrivial asymptotic power at this rate than the two other tests. The nontrivial asymptotic power for the  $Q$ -test against this sequence of local alternatives can also be observed for other values of the nominal size, as illustrated by the power-size plots for increasing sample sizes  $n$  shown in the right panel of Figure 3.

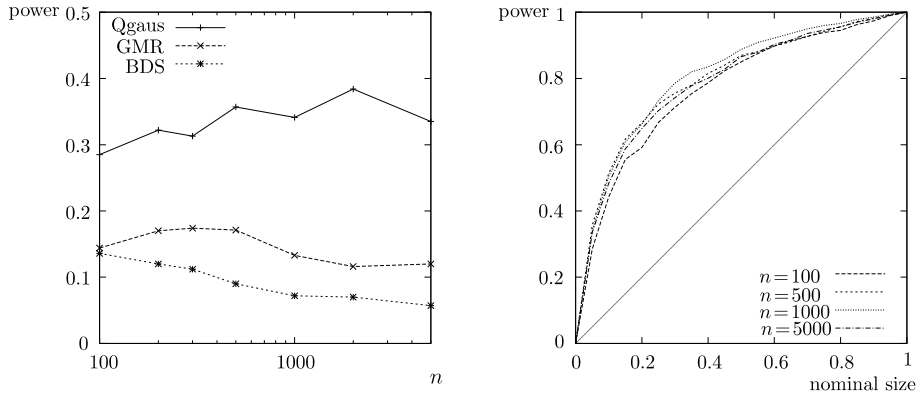


Figure 3. Observed power against local alternatives converging to the null at rate  $n^{-1/2}$ , as a function of sample size  $n = 100, \dots, 5,000$  at nominal size  $\alpha = 0.05$  (left panel); as a function of nominal size for the  $Q$ -test (right panel). Here lag  $\ell = 1$ , dimension  $m = 2$ , number of permutations  $B + 1 = 100$  ( $B + 1 = 200$  for BDS), number of simulations 1,000.

### 4.3. Application to estimated residuals

So far our theory and simulations were concerned only with the independence hypothesis for raw data. However, in practice the tests of independence are often used as specification tests while applied to the estimated residuals of some parametric model. Generally, estimated residuals are not independent and thus not exchangeable, even if they are based on i.i.d. innovations. The main question that determines the validity of the tests based on residuals is whether the dependence in the residuals introduced by parameter estimation affects the test statistic. A test employing parametrically estimated residuals will in general remain consistent if its rate is slower than the parametric rate, which is the case in the asymptotic test of Hong and White (2005) that is similar to the GMR test. Brock, Dechert, Scheinkman and LeBaron (1996) show that the presence of the estimated parameters does not affect the asymptotic distribution of their test statistic. Our simulations on estimated residuals show that the GMR and the BDS tests remain correct in terms of size. This is not the case, however, for the  $Q$ -test, at least for moderate sample sizes. In order to use the  $Q$ -test as a specification test on the estimated residuals we employ a parametric bootstrap (Efron (1979)). In this procedure we condition on  $k$  initial observations, where  $k$  is the order of the model, and the marginal distribution of the original residuals. The BDS and GMR permutation tests were applied directly to the residuals.

Table 2 shows rejection rates of the tests applied to estimated residuals of the AR(1) models and previously considered DGPs. Under the null, that is for DGP 4, the observed size of all tests is close to the nominal level 0.05. The power

of all tests drops compared to the tests of Subsection 4.1 based on raw data, which indicates that indeed some of the dependence structure is captured by the AR(1) model. The power of the  $Q$ -test on estimated residuals is comparable with that of the other tests, i.e., its power is lower for the nonlinear MA and bilinear processes (DGPs 1 – 3, 7), but it performs slightly better for the TAR model (DGP 10) and the logistic map (DGP 11).

Table 2. Observed rejection rates (size/power) for estimated residuals of the parametric AR (1) model (DGP 4), nominal size  $\alpha = 0.05$ , sample size  $n = 100$ , lag  $\ell = 1$ , dimension  $m = 2$ , number of permutations  $B + 1 = 100$  (200 for BDS) and number of simulations 1,000.

DGP	1	2	3	4	5	6	7	8	9	10	11	12	13
$Q$	0.61	0.63	0.12	0.04	0.13	0.08	0.12	0.21	0.11	0.37	0.79	0.66	0.54
BDS	0.68	0.70	0.27	0.06	0.06	0.06	0.21	0.47	0.21	0.13	0.52	0.26	0.13
GMR	0.70	0.72	0.15	0.04	0.11	0.07	0.12	0.24	0.07	0.33	0.68	0.63	0.42

## 5. Application to Financial Time Series

We consider an application to the Standard and Poor’s 500 Stock Index daily log-returns  $X_t = \ln(P_t/P_{t-1})$ , where  $P_t$  is the dividend-adjusted closing price index on day  $t$ , in the period 06/2001–05/2005 (source DATASTREAM). The sample was divided into two subsamples: Period 1 (06/2001–03/2003) and Period 2 (03/2003–05/2005), each having 500 observations. The sample division was made on the basis of visual inspection and basic statistics: Period 1 corresponds to a downward trend and exhibits strong volatility while Period 2 corresponds to an upward trend with moderate volatility. First, we test the geometric random walk hypothesis, which is equivalent to the null hypothesis of serial independence of the log-returns, using the  $Q$ -test for lags  $\ell = 1, \dots, 10$  and dimensions  $m = 2, 5$ .

The results (Table 3, columns “orig”) suggest that  $H_0$  is rejected for most of the lags for both periods. The evidence is stronger in the downward period and for the higher dimension ( $m = 5$ ). Next, we apply the test to the absolute values of the log-returns in a search for structure in volatility, and detect a stronger structure in volatility in the downward period (Table 3, columns “abs”). In an attempt to model the detected volatility structure we use an ARCH(1) specification and apply the  $Q$ -test to the absolute values of estimated residuals as a model specification test. Table 3 (columns “ARCH”) shows that the ARCH(1) filter is indeed able to capture the volatility structure for most of the lags and dimensions in the two periods.

Table 3. P-values based on the series of S&P 500 log-returns, their absolute values, and ARCH(1) filtered series for two periods. Here nominal size  $\alpha = 0.05$ , sample size  $n = 500$ , number of permutations  $B + 1 = 100$  and number of simulations 1,000.

		Period 1			Period 2		
$m$	$\ell$	orig	abs	ARCH	orig	abs	ARCH
2	1	0.09	0.13	0.24	0.01	0.06	0.15
2	2	0.04	0.01	0.18	0.14	0.26	0.39
2	3	0.15	0.01	0.14	0.93	0.45	0.58
2	5	0.13	0.01	0.27	0.64	0.40	0.56
2	10	0.36	0.06	0.41	0.04	0.73	0.16
5	1	0.02	0.01	0.03	0.02	0.08	0.21
5	2	0.01	0.01	0.05	0.08	0.03	0.09
5	3	0.05	0.01	0.25	0.54	0.04	0.19
5	5	0.01	0.01	0.14	0.04	0.60	0.19
5	10	0.81	0.13	0.16	0.02	0.05	0.06

## 6. Concluding Remarks

We introduced a new nonparametric test for serial independence based on quadratic forms. The test does not require the use of plug-in density estimators and remains consistent without letting the bandwidth diminish with sample size. We showed that the dependence measure used has desirable theoretical properties and several connections with other dependence measures. In particular we noticed that the test statistics are closely related to the statistics introduced by Rosenblatt (1975). Our findings imply that the latter statistics for fixed bandwidths have an interpretation as quadratic forms, so that they can be meaningfully used even if the underlying distributions are discontinuous.

We suggested a multiple bandwidth procedure to avoid the problem of optimal bandwidth selection while providing good power for various DGPs. Numerous simulations showed that the  $Q$ -test implemented on the basis of the exact permutation procedure has good finite sample performance against local and fixed alternatives in comparison with two other recent nonparametric tests: the BDS and GMR tests. The  $Q$ -test showed remarkably better power against TAR models. Further, we addressed the issue of using the  $Q$ -test as a parametric model specification test while applying it to residuals series and compared its performance in this situation with the BDS and the GMR tests. Finally, the test was applied to recent S&P 500 log-return series in downward- and upward-trend periods. The hypothesis of serial independence of the log-returns was rejected, with stronger rejection in the downward period. An application to residuals indicated that much of the structure in the volatility could be successfully accounted for by an ARCH(1) model.



## Acknowledgement

We express our gratitude to Prof. Esfandiar Maasoumi and Prof. Jeffrey Racine for kindly providing R code of their implementation of the GMR test. We wish to thank Prof. Jan de Gooijer and Dr. Kees Jan van Garderen for their detailed comments on earlier versions of this paper. We thank an associate editor and two anonymous referees for providing comments which greatly improved the paper.

## References

- Ahmad, I. A. and Li, Q. (1997). Testing independence by nonparametric kernel methods. *Statist. Probab. Lett.* **34**, 201-210.
- Anderson, N. H., Hall, P. and Titterton, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivariate Anal.* **50**, 41-54.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307-327.
- Brock, W., Dechert, W., Scheinkman, J. and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econom. Rev.* **15**, 197-235.
- Delgado, M. and Mora, J. (2000). A nonparametric test for serial independence of regression errors. *Biometrika* **87**, 228-234.
- Denker, M. and Keller, G. (1983). On  $U$ -statistics and  $v$ . Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete* **64**, 505-522.
- Diks, C. and Tong, H. (1999). A test for symmetry of multivariate probability distributions. *Biometrika* **86**, 605-614.
- Diks, C., van Zwet, W. R., Takens, F. and DeGoede, J. (1996). Detecting differences between delay vector distributions. *Phys. Rev. E* **53**, 2169-2176.
- Durbin, J. and Watson, G. (1950). Testing for serial correlation in least-squares regression. I. *Biometrika* **37**, 409-428.
- Durbin, J. and Watson, G. (1951). Testing for serial correlation in least-squares regression. II. *Biometrika* **38**, 159-177.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.
- Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987-1008.
- Feuerverger, A. (1993). A consistent test for bivariate dependence. *Internat. Statist. Rev.* **61**, 419-433.
- Good, P. (2000). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Series in Statistics. Springer-Verlag, New York.
- Granger, C. W. and Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht, Göttingen.
- Granger, C. W. and Lin, J. (1994). Using the mutual information coefficient to identify lags in nonlinear models. *J. Time Ser. Anal.* **15**, 371-384.
- Granger, C. W., Maasoumi, E. and Racine, J. (2004). A dependence metric for possibly nonlinear processes. *J. Time Ser. Anal.* **25**, 649-669.

- Granger, C. W. and Teräsvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Econom. Lett.* **62**, 161-165.
- Heer, G. R. (1991). Testing independence in high dimensions. *Statist. Probab. Lett.* **12**, 73-81.
- Hoeffding, W. (1948). A non-parametric test of independence. *Ann. Math. Statist.* **19**, 546-557.
- Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *Ann. Math. Statist.* **23**, 169-192.
- Hong, Y. (2000). Generalized spectral tests for serial dependence. *J. Roy. Statist. Soc. Ser. B* **62**, 557-574.
- Hong, Y. and White, H. (2005). Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* **73**, 837-901.
- Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599-631.
- Li, W. K. (2004). *Diagnostic Checks in Time Series*. Chapman and Hall/CRC.
- Marriott, F. (1979). Barnard's Monte Carlo tests: How many simulations? *Appl. Statist.* **28**, 75-77.
- Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any population. II. The correlation coefficient test. *Supplement to the J. Roy. Statist. Soc.* **4**, 225-232.
- Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3**, 1-14.
- Rosenblatt, M. and Wahlen, B. (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statist. Probab. Lett.* **15**, 245-252.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley, New York.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.
- Skaug, H. and Tjøstheim, D. (1993). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* **80**, 591-602.
- Székely, G. J. and Rizzo, M. L. (2005). A new test for multivariate normality. *J. Multivariate Anal.* **93**, 58-80.
- Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing* (Edited by C. H. Chen), 101-141. Sijthoff and Noordhoff, Amsterdam.
- Wolff, R. (1994). Independence in time series: another look at the BDS test. *Philos. Trans. R. Soc. Ser. A* **348**, 383-395.

Center for Nonlinear Dynamics in Economics and Finance, Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands.

E-mail: c.g.h.diks@uva.nl

School of Economics, University of New South Wales, Sydney, NSW 2052, Australia.

E-mail: v.panchenko@unsw.edu.au

(Received September 2005; accepted April 2006)

## NONPARAMETRIC TESTS FOR SERIAL INDEPENDENCE BASED ON QUADRATIC FORMS

Cees Diks and Valentyn Panchenko

*University of Amsterdam*

### Supplementary Material

#### A. Relation between $Q$ and the correlation coefficient

Our aim here is to find an analytic expression for the introduced distance measure, the quadratic form  $Q$  between the time series  $\{X_t\}$  of the above structure and a time series  $\{Y_t\}$  independently sampled from a multivariate normal distribution. The expression will be derived for the Gaussian product kernel  $K_h(x - y) = \prod_{i=1}^m \exp(-(x_i - y_i)^2 / (4h^2))$ . We consider a strictly stationary and weakly dependent time series  $\{X_t\}$  generated by a Gaussian process such that the  $m$ -dimensional delay vectors  $X_t^{m,\ell} = (X_t, X_{t+\ell}, \dots, X_{t+(m-1)\ell})'$  are multivariate normal random variables (standardised to unit variances) with correlation matrix  $\Omega$ . In the case of independence the correlation matrix reduces to the identity matrix. To simplify the integration we transform the multivariate normal pdf of the form  $f(x) = |\Omega|^{-1/2} (2\pi)^{-m/2} \exp(-(x'\Omega^{-1}x)/2)$  to  $z$  coordinates defined by  $z = Vx$ , where  $V$  is an orthogonal matrix and  $\Omega = VDV'$  by the spectral theorem, where  $D = \text{diag}(\eta_1^2, \dots, \eta_m^2)$ :  $f^*(z) = (2\pi)^{-m/2} \prod_{i=1}^m \eta_i^{-1} \exp(-z_i^2 / (2\eta_i^2))$ . The absolute value of the determinant of the Jacobian is one (property of an orthogonal matrix). Using the above transformation we can compute the elements of  $Q$ , letting  $f_0(\cdot)$  denote the product of marginal pdfs:

$$\begin{aligned}
 Q^{11} &= \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_h(r - s) f^*(r) f^*(s) dr ds = h^m \prod_{i=1}^m \frac{1}{\sqrt{h^2 + \eta_i^2}}, \\
 Q^{12} &= \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_h(r - s) f^*(r) f_0(s) dr ds = h^m \prod_{i=1}^m \frac{1}{\sqrt{h^2 + (\eta_i^2 + 1) / 2}}, \\
 Q^{22} &= \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_h(r - s) f_0(r) f_0(s) dr ds = h^m \prod_{i=1}^m \frac{1}{\sqrt{h^2 + 1}}.
 \end{aligned}$$

Combining terms we can express  $Q$  as a function of the eigenvalues  $\eta_i^2$  which are determined by the autocorrelations  $\rho_i$ , the bandwidth  $h$ , and the delay vector

dimension  $m$ :

$$Q = h^m \left( \prod_{i=1}^m \frac{1}{\sqrt{h^2 + \eta_i^2}} - 2 \prod_{i=1}^m \frac{1}{\sqrt{h^2 + \frac{(\eta_i^2 + 1)}{2}}} + \prod_{i=1}^m \frac{1}{\sqrt{h^2 + 1}} \right).$$

In the case of a bivariate standard normal distribution with a correlation coefficient  $\rho$ , the eigenvalues are simply expressed as  $\eta_1^2 = 1 + \rho$ ,  $\eta_2^2 = 1 - \rho$  and one obtains a direct correspondence between  $Q$  and  $\rho^2$ .

## B. Equivalence of $Q$ and quadratic distance

We establish the equivalence of the quadratic distance of Rosenblatt (1975)  $T = \int_{\mathbb{R}^m} (\hat{f}_1(x) - \hat{f}_2(x))^2 dx$  and the U-statistics estimator of quadratic form  $\hat{Q}$ . For simplicity we consider a Gaussian kernel for density estimation. Rewrite  $T$  explicitly in terms of the kernel density estimators:

$$T = \int_{\mathbb{R}^m} \left( \frac{1}{n_1} \sum_{t=1}^{n_1} \left( \frac{1}{\sqrt{2\pi}h} \right)^m e^{-\frac{\|X_t - x\|^2}{2h^2}} - \frac{1}{n_2} \sum_{s=1}^{n_2} \left( \frac{1}{\sqrt{2\pi}h} \right)^m e^{-\frac{\|Y_s - x\|^2}{2h^2}} \right)^2 dx.$$

Expanding the square, one arrives at the form  $T = T^{11} - 2T^{12} + T^{22}$ . For brevity we will derive the  $T^{11}$  term only, derivations for  $T^{12}$  and  $T^{22}$  being similar:

$$\begin{aligned} T^{11} &= \frac{1}{n_1^2} \left( \frac{1}{\sqrt{2\pi}h} \right)^{2m} \int_{\mathbb{R}^m} \sum_{t=1}^{n_1} \sum_{s=1}^{n_1} e^{-\frac{\|X_t - x\|^2 - \|X_s - x\|^2}{2h^2}} dx \\ &= \frac{1}{n_1^2} \left( \frac{1}{\sqrt{2\pi}h} \right)^{2m} \int_{\mathbb{R}^m} \sum_{t=1}^{n_1} \sum_{s=1}^{n_1} e^{-\frac{\|X_t - X_s\|^2}{4h^2}} e^{-\frac{-2\|x - (X_t + X_s)/2\|^2}{2h^2}} dx \\ &= \frac{1}{n_1^2} \left( \frac{1}{2\sqrt{\pi}h} \right)^m \sum_{t=1}^{n_1} \sum_{s=1}^{n_1} e^{-\frac{\|X_t - X_s\|^2}{4h^2}}. \end{aligned}$$

Above we used the Gaussian kernel factorisation that allows to reduce the analysis of the  $m$ -dimensional norm  $\|\cdot\|^2$  to one dimension  $(\cdot)^2$ . In this form  $T^{11}$  is exactly the same as the V-statistic estimator of  $Q^{11}$  times a factor  $(2\sqrt{\pi}h)^{-m}$  which does not depend on the data. Analogously, one can establish equivalence of  $T^{12}$ ,  $T^{22}$  and the V-statistic estimators of  $Q^{12}$  and  $Q^{22}$  respectively. Given the asymptotic equivalence of V-statistics and U-statistics we establish that  $T \approx (2\sqrt{\pi}h)^{-m} \hat{Q}$ .

## C. Performance for various kernels

The tables below report the rejection rates for five nonparametric tests for serial independence. The  $Q$ -columns correspond to the tests based on quadratic

forms with Gaussian, double exponential and Cauchy kernels respectively, and the remaining two columns to the BDS test and the GMR test, respectively. The nominal size of the tests was set to 0.05. We consider three lags  $\ell = 1, 2, 3$  for delay vector dimension  $m = 2$ , and only one lag  $\ell = 1$  for higher dimensions ( $m = 3, 4, 5, 10$ ). The bandwidth set  $H$  included  $d = 5$  different values in the range  $R = [0.5, 2.0]$  (after normalisation of the series to the unit variance). We used series of length  $n = 100$  ( $n = 50$  for DGP 6 and  $n = 20$  for DGP 11–13), the total number of permutations was set to  $B + 1 = 100$  (for BDS test  $B + 1 = 200$ ). The number of simulations was set to 1,000.

0.  $Y_t = \varepsilon_t$

1.  $Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$

$m$	$\ell$	$Q_g$	$Q_d$	$Q_c$	BDS	GMR	$Q_g$	$Q_d$	$Q_c$	BDS	GMR
2	1	0.06	0.07	0.06	0.05	0.05	0.71	0.81	0.76	0.76	0.78
2	2	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.07	0.06	0.04
3	1	0.05	0.06	0.06	0.05	0.06	0.71	0.78	0.77	0.68	0.57
5	1	0.05	0.06	0.05	0.05	0.05	0.57	0.64	0.62	0.46	0.39
10	1	0.05	0.04	0.05	0.06	0.05	0.32	0.37	0.35	0.22	0.20

2.  $Y_t = \varepsilon_t + 0.6\varepsilon_{t-1}^2 + 0.6\varepsilon_{t-2}^2$

3.  $Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}\varepsilon_{t-2}$

$m$	$\ell$	$Q_g$	$Q_d$	$Q_c$	BDS	GMR	$Q_g$	$Q_d$	$Q_c$	BDS	GMR
2	1	0.94	0.94	0.93	0.83	0.83	0.14	0.13	0.12	0.34	0.21
2	2	0.26	0.30	0.28	0.25	0.25	0.12	0.11	0.11	0.20	0.11
3	1	0.96	0.96	0.95	0.84	0.78	0.29	0.23	0.25	0.46	0.22
5	1	0.93	0.93	0.91	0.76	0.60	0.35	0.26	0.31	0.39	0.14
10	1	0.76	0.80	0.75	0.46	0.39	0.26	0.20	0.26	0.24	0.07

4.  $Y_t = 0.3Y_{t-1} + \varepsilon_t$

5.  $Y_t = 0.8|Y_{t-1}|^{0.5} + \varepsilon_t$

$m$	$\ell$	$Q_g$	$Q_d$	$Q_c$	BDS	GMR	$Q_g$	$Q_d$	$Q_c$	BDS	GMR
2	1	0.70	0.67	0.59	0.21	0.40	0.55	0.60	0.53	0.15	0.34
2	2	0.12	0.11	0.09	0.05	0.07	0.07	0.07	0.07	0.05	0.06
3	1	0.68	0.64	0.56	0.16	0.31	0.53	0.57	0.50	0.11	0.25
5	1	0.61	0.58	0.46	0.13	0.20	0.43	0.48	0.40	0.09	0.18
10	1	0.45	0.45	0.32	0.07	0.15	0.29	0.35	0.25	0.06	0.12

6.  $Y_t = \text{sign}(Y_{t-1}) + \varepsilon_t$

7.  $Y_t = 0.6\varepsilon_{t-1}Y_{t-2} + \varepsilon_t$

$m$	$\ell$	$Q_g$	$Q_d$	$Q_c$	BDS	GMR	$Q_g$	$Q_d$	$Q_c$	BDS	GMR
2	1	0.98	0.98	0.98	0.77	0.90	0.18	0.17	0.16	0.32	0.17
2	2	0.70	0.72	0.68	0.33	0.37	0.20	0.18	0.18	0.41	0.22
3	1	0.98	0.98	0.97	0.75	0.86	0.39	0.31	0.33	0.50	0.26
5	1	0.96	0.97	0.95	0.64	0.74	0.48	0.38	0.43	0.51	0.21
10	1	0.90	0.92	0.87	0.40	0.54	0.38	0.33	0.35	0.34	0.14

8.  $Y_t = \varepsilon_t \sqrt{1 + 0.4Y_{t-1}^2}$

9.  $Y_t = \sqrt{h_t} \varepsilon_t,$   
 $h_t = 0.01 + 0.80h_{t-1} + 0.15Y_{t-1}^2$

$m$	$\ell$	$Q_g$	$Q_d$	$Q_c$	BDS	GMR	$Q_g$	$Q_d$	$Q_c$	BDS	GMR
2	1	0.25	0.22	0.23	0.55	0.32	0.13	0.12	0.12	0.26	0.13
2	2	0.07	0.07	0.06	0.12	0.09	0.12	0.12	0.11	0.23	0.11
3	1	0.24	0.22	0.24	0.51	0.26	0.18	0.15	0.16	0.35	0.15
5	1	0.24	0.20	0.21	0.39	0.18	0.26	0.23	0.24	0.43	0.18
10	1	0.20	0.16	0.18	0.24	0.11	0.38	0.32	0.36	0.48	0.17

10.  $Y_t = \begin{cases} -0.5Y_{t-1} + \varepsilon_t, & Y_{t-1} < 1 \\ 0.4Y_{t-1} + \varepsilon_t, & \text{else} \end{cases}$

11.  $Y_t = 4Y_{t-1}(1 - Y_{t-1}), 0 < Y_t < 1$

$m$	$\ell$	$Q_g$	$Q_d$	$Q_c$	BDS	GMR	$Q_g$	$Q_d$	$Q_c$	BDS	GMR
2	1	0.91	0.94	0.90	0.07	0.49	0.98	0.97	1.00	0.85	0.96
2	2	0.10	0.09	0.10	0.06	0.05	0.08	0.12	0.23	0.16	0.74
3	1	0.87	0.91	0.87	0.06	0.34	0.90	0.92	1.00	0.71	0.95
5	1	0.77	0.83	0.76	0.04	0.23	0.54	0.59	0.79	0.35	0.88
10	1	0.49	0.61	0.47	0.03	0.14	0.17	0.24	0.31	0.11	0.59

12.  $Y_t = 1 + 0.3Y_{t-2} - 1.4Y_{t-1}^2$

13.  $Y_t = Z_t + \sigma\varepsilon_t,$   
 $Z_t = 1 + 0.3Z_{t-2} - 1.4Z_{t-1}^2$

$m$	$\ell$	$Q_g$	$Q_d$	$Q_c$	BDS	GMR	$Q_g$	$Q_d$	$Q_c$	BDS	GMR
2	1	0.99	0.99	1.00	0.46	0.99	0.93	0.91	0.95	0.30	0.59
2	2	0.16	0.22	0.50	0.13	0.55	0.15	0.15	0.26	0.06	0.13
3	1	0.97	0.97	1.00	0.46	0.96	0.83	0.84	0.91	0.22	0.41
5	1	0.86	0.89	0.96	0.35	0.87	0.74	0.73	0.81	0.10	0.25
10	1	0.40	0.39	0.51	0.12	0.46	0.37	0.34	0.45	0.39	0.12