Nonparametric Tests for Serial Independence Based on Quadratic Forms

Cees Diks & Valentyn Panchenko

Center for Nonlinear Dynamics in Economics and Finance, Department of Economics,

University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

July 17, 2005

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. 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. It is beyond the scope of this paper to provide a complete summary of the nonparametric dependence measures considered in the litereature. The interested reader is referred to Tjøstheim (1996) for a comprehensive review.

Consider a strictly stationary and mixing real-valued time series process $\{X_t\}$, $t \in \mathbb{Z}$. We introduce the m-dimensional lag ℓ delay vectors $X_t^{m,\ell} = (X_t, X_{t+\ell}, X_{t+2\ell} \ldots, X_{t+(m-1)\ell})$ and denote the time invariant probability measure of $X_t^{m,\ell}$ by ν_m, suppressing the dependence on ℓ, 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 identially distributed (i.i.d.). Under the null hypothesis the m-dimensional delay vector measure ν_m is equal to the product measure $\nu_1^m = \nu_1 \times \cdots \times \nu_1$ (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 $x = (x_1, \ldots, x_m)$, and the null hypothesis can be expressed as $f_m(x) = f_1(x_1) \times \cdots \times f_1(x_m)$.

The idea to measure the divergence between the joint measure ν^m and the product measure ν_1^m dates back to Hoeffding (1948). Recently, information theoretic measures of divergence such as the Kullback-Leibler information criterion 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 recognized 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 $T = \int(\widehat{f}_1 - \widehat{f}_2)^2$, where $\widehat{f}_i$ 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 to produce 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 which are closely related to the statistic $T = \int(\widehat{f}_1 - \widehat{f}_2)^2$. However, it becomes apparent that the bandwidth plays an entirely different role than in nonparametric density estimation, where it controls the trade-off between the bias and variance of the density estimators $\widehat{f}_i$. 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, parameterized 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.

Using the results by Denker and Keller (1983) it will be shown that under some mild regularity conditions such as strict stationarity and mixing conditions, the proposed tests for serial independence are consistent against all fixed alternatives. In contrast to some other tests based on quadratric 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 invesigates 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. Section 5 investigates how the test may be applied to estimated residuals. After an application to the log-return series of S&P 500 stock index in section 6, section 7 summarises and concludes.

2 Quadratic forms and their estimators

In this section we briefly review the distance notions between probability measures in $\mathbb{R}^m$ which will serve as the divergence between the joint probability measure ν_m and its counterpart ν_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 Diks and Tong (1999). For two m-dimensional probability measures μ_1 and μ_2, consider a quadratic form of the type:

$$Q = \|\mu_1 - \mu_2\|^2 = \langle \mu_1 - \mu_2, \mu_1 - \mu_2 \rangle$$

where $\langle \mu_1, \mu_2 \rangle = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K(x_1, x_2)\mu_1(dx_1)\mu_2(dx_2)$ is a bilinear form. Whenever $K(\cdot, \cdot)$ 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 factorize as $K_h(x_1, x_2) = \prod_{i=1}^m \kappa((x_{1i} - x_{2i})/h)$ where $\kappa(\cdot)$ is a one-dimensional kernel function, which is symmetric around zero, and where 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

$$Q = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_h(x_1, x_2)(\mu_1 - \mu_2)(dx_1)(\mu_1 - \mu_2)(dx_2) = \int_{\mathbb{R}^m} \tilde{K}_h(k)|(\tilde{\mu}_1 - \tilde{\mu}_2)(k)|^2 dk,$$

where $\tilde{K}_h(k)$ is the Fourier transform of the kernel, $\tilde{\mu}_i$ the characteristic function of μ_i, and $|\cdot|$ the modulus. It follows that if $K_h(x_1, x_2)$ is positive definite, $\tilde{K}_h(k)$ is a real-valued positive function which does not vanish on any interval, and $Q = 0$ if and only if $\mu_1 = \mu_2$ almost everywhere 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 summarized by Granger, Maasoumi, and Racine (2004). It is well-defined for both continuous and discrete random variables. It is nonnegative, equal to zero only in the case of independence, and can be related to the correlation coefficient ρ in the case of a bivariate gaussian distribution, as shown in Appendix A. Since $\langle \cdot, \cdot \rangle$ is an inner product, the distance $Q^{\frac{1}{2}}$ is a metric with usual properties of a metric such as the triangular inequality. Although Q is not invariant under monotonic nondecreasing transformations, if desired, invariance of estimators can always be achieved by transforming the data to a known marginal distribution, e.g. by using the empirical probability integral transform. Moreover, in Appendix B we establish the equivalency between the quadratic form Q and the quadratic measure of Rosenblatt (1975).

Using the bilinearity of the inner product Q can be written as $Q = Q_{11} - 2Q_{12} + Q_{22}$, where $Q_{ij} = \langle \mu_i, \mu_j \rangle = E\left[K(Z, Z')\right]$, with $Z \sim \mu_i$ and $Z' \sim \mu_j$ independent. The fact that the 'inner products' Q_{ij} between μ_i and μ_j can be expressed in terms of averages of a kernel function, suggests the use of U- or V-statistics for their estimation.

Given an observed sequence $\{X_t\}_{t=1}^{T}$, from which one can construct $n = T - (m-1)\ell$ delay vectors $X_t^{m,\ell}$, $t = 1, \ldots, n$ of dimension m, for the first term Q_{11} this gives the U-statistic estimator

$$\widehat{Q}_{11} = \frac{2}{n(n-1)} \sum_{t=2}^{n} \sum_{t'=1}^{t-1} K_h(X_t^{m,\ell} - X_{t'}^{m,\ell}) = \frac{2}{n(n-1)} \sum_{t=2}^{n} \sum_{t'=1}^{t-1} \prod_{k=0}^{m-1} \kappa\left((X_{t+k\ell} - X_{t'+k\ell})/h\right).$$

For the bounded kernel functions considered here, it follows from 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 $\widehat{C}_h(x) = \frac{1}{n}\sum_{t=1}^{n} \kappa\left((x - X_t)/h\right)$ for $E[\kappa((x-X)/h)]$ and use this to obtain consistent estimators for Q_{12} and Q_{22} which can be written as functionals of $E[\kappa((x-X)/h)]$:

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

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 $\kappa(x) = I_{[0,h)}(x)$, which is 1 if $x \in [0,h)$ 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 for the U-statistics it is possible to develop asymptotic theory for the functional Q. However, as reported by Skaug and Tjøstheim (1993), Granger, Maasoumi and Racine (2004) and Hong and White (2005) in similar testing contexts, the asymptotic theory provides a rather poor finite sample approximation to the null distribution of the test statistic 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 for any sample size n it is exact, i.e. the rejection rate under the null hypothesis is equal to the nominal size α. 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 large values of Q indicate deviations from the null distribution, a test based on this squared distance would be to 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 values of $\widehat{Q}_i$, $i = 0, \dots, B$, computed using the original data and B permutations, are exchangeable. An exact p-value (in that it is uniformly distributed on $\frac{1}{B+1}, \frac{2}{B+1}, \dots, 1$ under the null) is calculated as

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

where $I(\cdot)$ denotes indicator function that takes value of 1 if the condition in brackets is true and 0 otherwise. Let $R = \sum_{i=0}^{B} I(\widehat{Q}_i = \widehat{Q}_0) \geq 1$ denote the number of ties plus one. In case $R = 1$, $L = 1$, while for $R > 1$, for L we take a random variable, uniformly distributed on $1, \ldots, R$. 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-α 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, section 3.3 describes a method for determining a single p-value over a range of different bandwidths. However, we first motivate the multiple bandwidths procedure by presenting the results of some bandwidth-related simulations.

3.2 Bandwidth-related simulations

Hereafter we refer to the bandwidth that yields the highest empirical power for a fixed size α as the optimal bandwidth h^*. We investigate the dependence of the optimal bandwidth on the 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 30 different bandwidth values h_k ranging from 0.01 to 10.00, equidistant on a logarithmic scale:

$$h_k \quad = \quad h_{\max}(h_{\min}/h_{\max})^{\frac{K-k}{K-1}}, \qquad k = 1, \ldots, K. \tag{2}$$

The number of permutation 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 report results only for the gaussian kernel.

Figure 1 shows the power as a function of the bandwidth for series of various lengths n, (left, DGP 1, $m = 2$, $l = 1$), for various dimensions m, (middle, DGP 10, $n = 100$, $l = 1$) and for various DGPs, (right, $n = 100$, $m = 2$, $l = 1$). The left plot 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 joint delay vector measure ν_m and product measure ν_1^m, and if this length scale does not depend on n the optimal bandwidth is asymptotically independent of n. Analytical support for a fixed optimal bandwidth was reported by Anderson, Hall, and Titterington (1994) in a two-sample test based on a statistic of the type $T = \int(\widehat{f}_1 - \widehat{f}_2)^2$. In Appendix B we show that this functional can in fact be interpreted as a quadratic form.

However, the right panel of Figure 1 illustrates that the optimal bandwidth h^* depends on the particular DGP, e.g. for DGP 1, the MA(1) process, $h^* = 0.7$, and for DGP 7, the bilinear process $h^* = 1.2$. This suggests that using a single bandwidth value in a practical situation, when the underlying DGP is not known, may not be optimal.

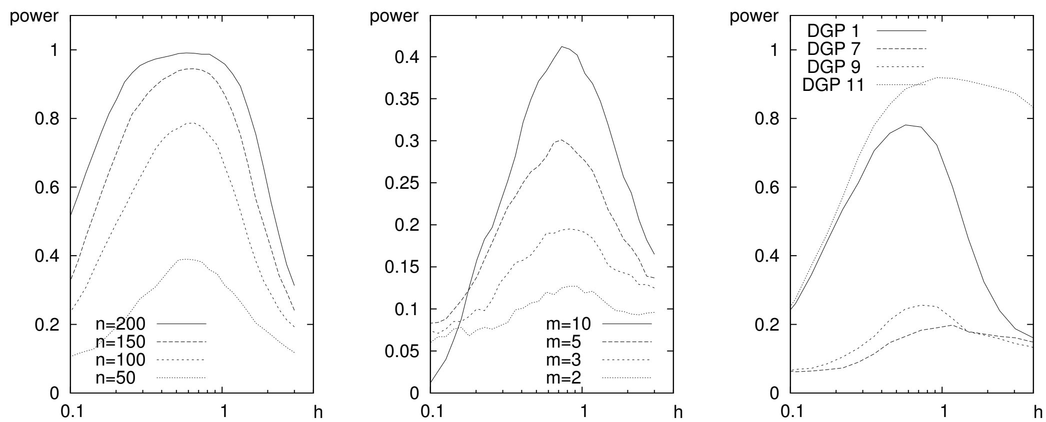

Figure 1: *Power as a function of bandwidth h for various series lengths n, DGP 1, dimension $m = 2$ (left), various dimensions m, DGP 10, $n = 100$ (middle) and various DGPs, $n = 100$, $m = 2$ (right). Lag $l = 1$ in all cases, nominal size $\alpha = 0.05$, number of permutations $B + 1 = 100$, number of simulations* 1 000.

3.3 Multiple bandwidth procedure

Motivated by the findings of the previous subsection, 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 bandwidth. 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 summarized as follows:

1. Calculate the vector of $\widehat{Q}_{h,0}$-values for a range of bandwidths: $h \in H = \{h_1, \ldots, h_K\}$. We define h on a geometric grid as in Eq. 2.

2. Randomly permute the data and calculate a bootstrap vector $\widehat{Q}_{h,1}$. Repeat this B

times, to obtain $\widehat{Q}_{h,i}$ for $h \in H$, and $i = 1, \ldots, B$.

3. Transform $\widehat{Q}_{h,i}$ into a p-value: $\widehat{p}_{h,i} = [\sum_{j=0}^{B} I(\widehat{Q}_{h,j} > \widehat{Q}_{h,i}) + L]/(B+1)$, with L defined similarly to Eq. 1.

4. Select the smallest p-value among all bandwidths and call it $\widehat{T}_i$: $\widehat{T}_i = \inf_{h \in H} \widehat{p}_{h,i}$.

5. Calculate an overall p-value on the basis of the rank of $\widehat{T}_0$ among the $\widehat{T}_i$ $(i = 0, \ldots, B)$, i.e. $\widehat{p} = [\sum_{j=0}^{B}(\widehat{T}_i < \widehat{T}_0) + L]/(B+1)$ using the ties randomization procedure as in Eq. 1.

In step 3 we pretend each of the permuted series to be the originally observed series and determine the corresponding p-values $\widehat{p}_{h,i}$ that would have been obtained for series i for each of the different bandwidths. In step 4, for each series the smallest p-values over different the bandwidths is selected (denoted by $\widehat{T}_i$, $i = 0, \ldots, B$). We finally use the exchangeability of the B series under the null to calculate an overall p-value by establishing the significance of $\widehat{T}_0$ for the actually observed data (step 5). As in the single bandwidth case, the multiple bandwidth procedure yields an exact α-level test which rejects the null hypothesis if $\widehat{p} \leq \alpha$. The power of this multiple-bandwidth procedure depends on the width of the region $R = [h_{\max}, h_{\min}]$, the number K 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 K chosen in R is important for the power. Taking K too small we risk 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 $R = [0.5, 2]$ which includes h^* for all considered DGPs. For this region reasonable power is achieved using $K = 5$.

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 the this procedure is more sensitive to the number of permutations B than for the single bandwidth procedure. The reason for this is that the $\widehat{T}_i$ are discrete multiples of $1/(B+1)$, which for small B leads to many identical$\widehat{T}_i$-values (ties) which reduces the power. We find that for the considered region $R = [0.5, 2]$ with $K = 5$, 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) suggests that under certain conditions the random critical value of the permutation test converges to a constant as the number of permutation $B \to \infty$ as $n \to \infty$. Relying on this fact Hoeffding investigated the large-sample power properties of the permutation tests based on a relatively simple test statistic and demonstrated that under general conditions the permutations 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:

DGP 0. $\quad y_t = \varepsilon_t$

DGP 1. $\quad y_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$

DGP 2. $\quad y_t = \varepsilon_t + 0.6\varepsilon_{t-1}^2 + 0.6\varepsilon_{t-2}^2$

DGP 3. $\quad y_t = \varepsilon_t + 0.8\varepsilon_{t-1}\varepsilon_{t-2}$

DGP 4. $\quad y_t = 0.3y_{t-1} + \varepsilon_t$

DGP 5. $\quad y_t = 0.8|y_{t-1}|^{0.5} + \varepsilon_t$

DGP 6. $\quad y_t = \text{sign}(y_{t-1}) + \varepsilon_t$

DGP 7. $\quad y_t = 0.6\varepsilon_{t-1}y_{t-2} + \varepsilon_t$

DGP 8. $\quad y_t = 4y_{t-1}(1 - y_{t-1}) \qquad 0 < y_t < 1$

DGP 9. $\quad y_t = \sqrt{h_t}\varepsilon_t, \qquad h_t = 1 + 0.4y_{t-1}^2$

DGP 10. $\quad y_t = \sqrt{h_t}\varepsilon_t, \qquad h_t = 0.01 + 0.80h_{t-1} + 0.15y_{t-1}^2$

DGP 11. $\quad y_t = \begin{cases} -0.5y_{t-1} + \varepsilon_t, & y_{t-1} < 1 \\ 0.4y_{t-1} + \varepsilon_t, & \text{else} \end{cases}$

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 used as a diagnostic tool for the empirical size of the suggested test. 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 lags. DGP 4 is a linear AR(1) process. DGP $5-6$ 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). DGP 8 is the logistic map generating deterministic chaotic time series. DGP $9-10$ are instances of the 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 11 is a TAR(1) process proposed by Tong (1978). We used series of length $n = 100$ (except $n = 50$ for DGP 6 and $n = 20$ for DGP 8), and the total number of bootstrap replications (permutations) was set to $B + 1 = 100$, including the original series. The bandwidth set H included $K = 5$ different values in the range $R = [0.5, 2.00]$ (after normalising the series to unit variance). The three different kernels mentioned earlier were used for comparison: gaussian, double exponential and Cauchy. We consider different lags $l = 1, 2, 3$ for delay vector dimension $m = 2$ and extend the delay vector dimension to $m = 3, 4, 5, 10$ for lag $l = 1$. All tests were conducted at a nominal size of $\alpha = 0.05$. 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 DGP 8 the rejection rate was smaller than the nominal size while using one-sided test. Therefore, we implement it as a two-sided test. To make the BDS test comparable with the Q-test we apply similar multiple bandwidth permutation procedure and double $B + 1 = 200$ to take into account the two-sizedness. The bandwidth range $R = [0.5, 2.0]$, which is typical for the BDS test, coincides with the one we use in the Q-test. We set $K = 5$ 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.

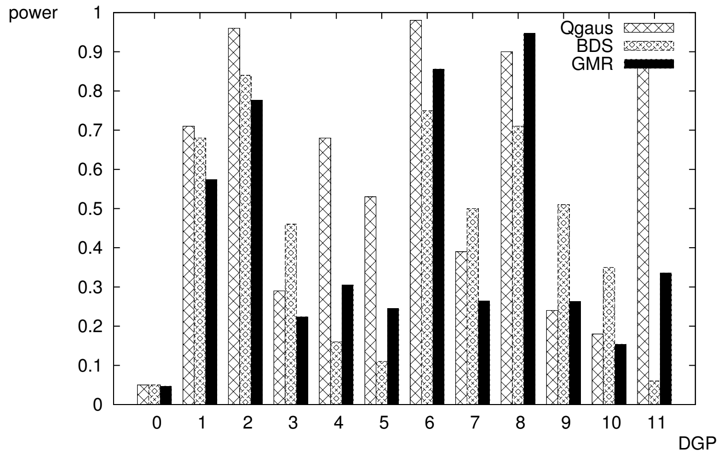

Figure 2: *Rejection rates against various DGPs. Nominal size* $\alpha = 0.05$*, sample size* $n = 100$*, lag* $l = 1$*, dimension* $m = 3$*, number of permutations* $B + 1 = 100$ *(*200 *for BDS), number of simulations* 1 000*.*

3.4.4) to obtain an optimal bandwidth (determining separate optimal bandwidth values under the null and an alternative), we do not undertake any further bandwidth adaptation. For embedding dimensions higher than two we use their "portmanteau" version of the test.

Figure 2 reports the observed rejection rates (at size $\alpha = 0.05$, $l = 1$, $m = 3$) for the considered processes for the introduced Q-test based on the gaussian kernel, the BDS test and the GMR test. For the numerical values and extended results (higher lags l and dimensions m) of these tests and the Q-tests based on other kernels we refer the interested reader to Appendix C. 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 most cases outperforms them, i.e. DGP $1 - 2, 4 - 6, 11$, nonlinear MA(1)-MA(2), linear, fractional and sign function AR(1), and TAR(1). In absolute terms the power of the Q-test is smaller for DGP $3, 7, 8$, i.e. nonlinear MA(2), bilinear process and logistic map, but still comparable to that obtained by the best performing test (for a particular DGP). In comparison with BDS test, the Q-test shows less power for DGP $9, 10$, ARCH(1) and GARCH$(1, 1)$. The GMR test behaves similar as 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. Therefore, we proceed further with the analysis based on the gaussian kernel only.

DGP 9 and its generalization DGP 10 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 DGP 10 is of special interest in financial econometric 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 DGP 10 we can adopt a semi-parametric approach and before testing transform the data to their absolute values. Table 1 shows the rejection rates for test based on the DGP 10 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. Transformation to the absolute values makes large, but differently signed points locally close to each other, 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 the case of 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

	Qgaus		BDS		GMR	
m	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

Table 1: *Rejection rates against DGP* 10 *after/before transforming the data to absolute values, nominal size* $\alpha = 0.05$*, sample size* $n = 100$*, lag* $l = 1$*, number of replication* $B + 1 = 100$ *(*200 *for BDS) and number of simulations* 1 000*.*

local alternative reduces at the rate $n^{-1/2}h^{-1/2}$ with $h \to 0$, which is 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 diminishing with the sample size. Therefore, we expect that the test has nontrivial asymptotic power at 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:

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

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

In the simulations we use an MA(1) process $y_t = \varepsilon_t + a_n\varepsilon_{t-1}$ with ε_t independent iden-

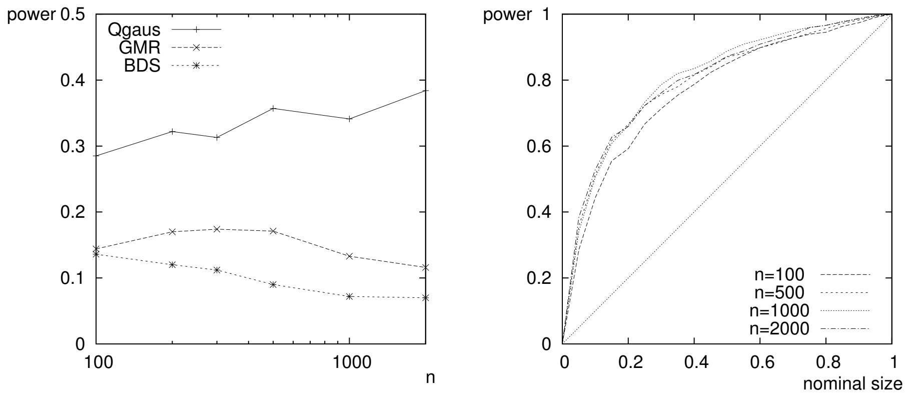

Figure 3: *Rejection rates against local alternative converging to the null at rate* $n^{-1/2}$*: (left) as function of sample size* $n = 100, \ldots, 2000$ *at nominal size* $\alpha = 0.05$*; (right) as a function of nominal size for the* Q*-test. Lag* $l = 1$*, dimension* $m = 2$*, number of permutations* $B + 1 = 100$ *(*$B + 1 = 200$ *for BDS), number of simulations* 1 000.

tically distributed (i.i.d.) $N(0, 1)$, the pdf of which can be represented in the form (3) with $q_j(y_t, y_{t+j}) = y_t y_{t+j}$. Figure 3 (left) shows the rejection rates (powers) of the considered test against a sequence of local alternatives which converges to the null at the usual parametric rate $a_n = Cn^{-1/2}$, where $C = 2$ and $n = 100, \ldots, 2000$. A horizontal line in the graph would indicate the parametric rate. After an initial transient period, $n < 300$, the rate for the GMR test becomes slower than parametric. The BDS test asymptotically indeed approaches the parametric rate, as well as the Q-test which appears to have substantial nontrivial asymptotic power at the parametric rate $n^{-1/2}$. The latter is also illustrated by the power-size plots for various values of the sample size n against the same local alternative (Figure 3, right).

5 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 which determines the validity of the tests based on residuals is whether the dependence in the residuals introduced by parameter estimation affects the test statistic. In the case of parametric regression parameters are estimated at the rate $n^{-1/2}$. A test employing parametrically estimated residuals will in general remain consistent if its rate is slower than the parametric, which is the case in the asymptotic test of Hong and White (2005) 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. A possible reason for this is that the estimate of Q has a higher than parametric rate of convergence under the null. In many cases we found that the Q-test statistic gets negatively biased, resulting in a very conservative permutation test. The bias does not reduce with increasing sample size. In order to use the Q-statistics as a specification test on the estimated residuals we employ a parametric bootstrap (Efron, 1979). The procedure is implemented as follows:

1. Specify a parametric model, estimate a vector of parameters $\widehat{\beta}_0$ and compute residuals $\widehat{\varepsilon}_0$.
2. Standardize the residuals to the unit variance and compute $\widehat{Q}_{h,0}$.
3. Use $\widehat{\beta}_0$ in conjunction with the specified parametric model and permuted residuals $\widehat{\varepsilon}_0$

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

Table 2: *Rejection rates based on the series of estimated residuals of the parametric AR* (1) *model (DGP* 4) *, nominal size* $\alpha = 0.05$*, sample size* $n = 100$*, lag* $l = 1$*, dimension* $m = 2$*, number of replication* $B + 1 = 100$ *(*200 *for BDS) and number of simulations* 1 000*.*

as innovations to obtain bootstrapped data.

4. Reestimate the model on the bootstrapped data and compute standardized residuals $\widehat{\varepsilon}_i$.

5. Compute $\widehat{Q}_{h,i}$ on the basis of $\widehat{\varepsilon}_i$. Repeat the steps $(3-5)$ $B+1$ times using the multiple bandwidth procedure of Subsection 3.3 to compute an overall $\widehat{p}$-value.

In step 3 we condition on a number of initial original observations, equal to the order of the model, and the marginal distribution of the original residuals. In contrast to the permutation procedure the parametric bootstrap does not yield an exact test for finite sample sizes. 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 MA and bilinear processes, but it performs slightly better for the logistic map and the TAR (1) model.

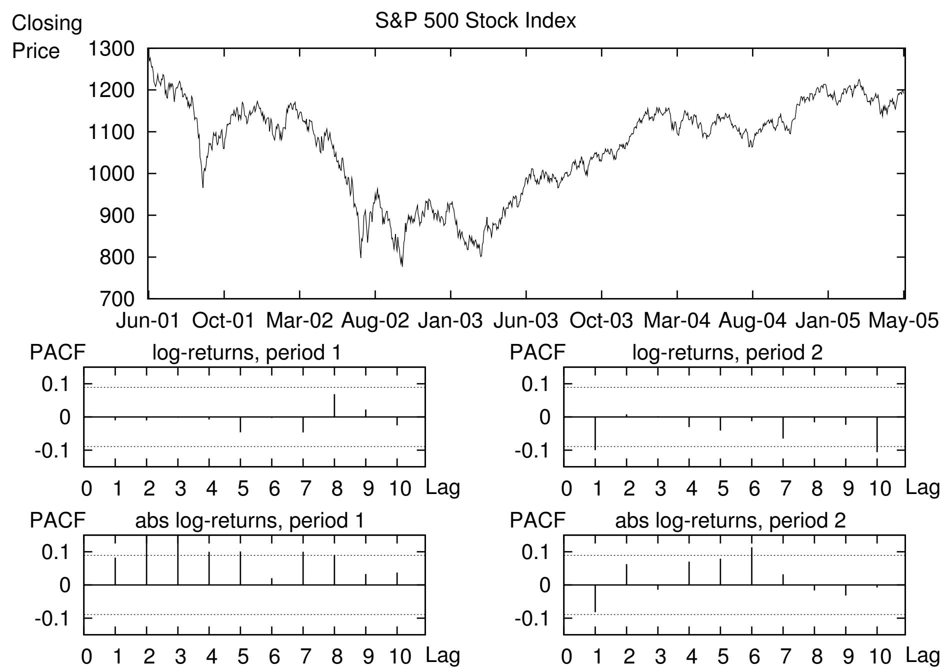

Figure 4: *Daily series of the S&P500 Stock Index and PACF plots of log-return and absolute log-return series ($n = 500$) for the two periods.*

6 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. Figure 4 shows the daily time series in levels of the S&P500 Stock Index as well as the partial autocorrelation function (PACF) plots of the log-returns and absolute log-returns series for the two periods. 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 for

		Period 1			Period 2		
m	l	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

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

a geometric random walk hypothesis, which is equivalent to the null hypothesis of serial independence of the log-returns, using the Q-test for lags $l = 1, \ldots, 10$ 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 search for a structure in volatility and detect a stronger structure in volatility in the downward period (Table 3, columns "abs"). Comparing the results of the Q-test ($m = 2$) with the PACFs in Figure 4 we notice that both tests reject the null at the same lags. This shows that the fully nonparametric Q-test is able to detect the same structure as a commonly used parametric test. In an attempt to model the detected volatility structure we use an ARCH(1) specification and apply the Q-test on 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 embedding dimensions in the two periods.

7 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 that the latter statistics for fixed bandwidths have an interpretation as quadratic forms, 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 recent nonparametric tests such as BDS and GMR. 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 the recent S&P 500 log-return series in downward- and upward-trend periods. The serial independence of the log-return was rejected, with stronger rejection in the downward period. An application to residuals indicated that much of the structure in the volatility that was successfully accounted for by an ARCH(1) model.

Acknowledgements

We express our gratitude to prof. Esfandiar Maasoumi and prof. Jeffrey Racine for kindly providing R code of their implementation of the GMR test. The usual disclaimers apply.

Appendix A. Relation between correlation coefficient and quadratic form Q

Consider a strictly stationary and mixing time series $\{X_i\}$ generated by a gaussian process such that the m-dimensional delay vectors $\boldsymbol{X}_i = (X_i, X_{i+1}, \ldots, X_{i+m-1})$ are multivariate normal random variables (standardized to unit variances) with correlation matrix Ω. In the case of independence the correlation matrix reduces to the identity matrix. Our aim here is to find an analytic expression for the introduced distance measure, the quadratic form Q between the time series $\{X_i\}$ of the above structure and a time series $\{Y_i\}$ independently sampled from a standard normal distribution. The expression will be derived for the gaussian kernel $K_h(x, y) = \prod_{i=1}^{m} \exp\left(-(x_i - y_i)^2/(4h^2)\right)$. To simplify the integration we transform the multivariate normal pdf of the form $f(\boldsymbol{x}) = |\Omega|^{-1/2}\,(2\pi)^{-m/2}\exp(-\frac{1}{2}(\boldsymbol{x}'\Omega^{-1}\boldsymbol{x}))$ to $\boldsymbol{z}$ coordinates defined by $\boldsymbol{z} = V\boldsymbol{x}$, where V is an orthogonal matrix and $\Omega = VDV'$ by the spectral theorem, where the diagonal matrix D contains eigenvalues $\lambda_1, \ldots, \lambda_m$ on the diagonal (denote $\eta_i^2 = \lambda_i$): $f^*(\boldsymbol{z}) = (2\pi)^{-(m)/2} \prod_{i=0}^{m} \eta_i^{-1} \exp(-z_i^2/(2\eta_i^2))$. The absolute value 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:

$$
\begin{aligned}
Q_{11} &= \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} K\left(\boldsymbol{z}_1,\boldsymbol{z}_2\right) f^*(\boldsymbol{z}_1) f^*(\boldsymbol{z}_2) d\boldsymbol{z}_1 d\boldsymbol{z}_2 = h^m \prod_{i=1}^{m} \frac{1}{\sqrt{h^2+\eta_i^2}},\\
Q_{12} &= \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} K\left(\boldsymbol{z}_1,\boldsymbol{y}\right) f^*(\boldsymbol{z}_1) f_0(\boldsymbol{y}) d\boldsymbol{z}_1 d\boldsymbol{y} = h^m \prod_{i=1}^{m} \frac{1}{\sqrt{h^2+\left(\eta_i^2+1\right)/2}},\\
Q_{22} &= \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} K\left(\boldsymbol{y}_1,\boldsymbol{y}_2\right) f_0(\boldsymbol{y}_1) f_0(\boldsymbol{y}_2) d\boldsymbol{y}_1 d\boldsymbol{y}_2 = h^m \prod_{i=1}^{m} \frac{1}{\sqrt{h^2+1}}.
\end{aligned}
$$

Combining terms we can express Q as a function of eigenvalues η_i^2 which are determined by the autocorrelations ρ_i, bandwidth h, and the delay vector dimension m

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

In the case of a bivariate standard normal distribution with a correlation coefficient ρ, 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 ρ^2.

Appendix B. Equivalence of quadratic form Q and quadratic distance

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

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

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

$$\begin{aligned} T_{11} &= \frac{1}{n_1^2}\left(\frac{1}{\sqrt{2\pi}h}\right)^{2m} \int_{\mathbb{R}^m} \sum_{t=1}^{n_1}\sum_{t'=1}^{n_1} e^{\frac{-\|X_{1t}-x\|^2 - \left\|X_{1t'}-x\right\|^2}{2h^2}} dx \\ &= \left(\frac{1}{\sqrt{2\pi}h}\right)^{2m} \int_{\mathbb{R}^m} e^{\frac{-2\left\|x-(X_{1t}+X_{1t'})/2\right\|^2}{2h^2}} \frac{1}{n_1^2}\sum_{t=1}^{n_1}\sum_{t'=1}^{n_1} e^{\frac{-\left\|X_{1t}-X_{1t'}\right\|^2}{4h^2}} dx \\ &= \left(\frac{1}{\sqrt{2\pi}h}\frac{1}{\sqrt{2}}\right)^{m} \frac{1}{n_1^2}\sum_{t=1}^{n_1}\sum_{t'=1}^{n_1} e^{\frac{-\left\|X_{1t}-X_{1t'}\right\|^2}{4h^2}} dx. \end{aligned}$$

Above we used the gaussian kernel factorization 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 $\left(\frac{1}{\sqrt{2\pi}h}\frac{1}{\sqrt{2}}\right)^m$ 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 $T \approx \left(\frac{1}{\sqrt{2\pi}h}\frac{1}{\sqrt{2}}\right)^m \hat{Q}$.

Appendix C. Finite sample performance of nonparametric tests for serial Independence

The tables below report the rejection rates for five nonparametric tests for serial independence. The first three are the tests based on quadratic forms with gaussian, double exponential and Cauchy kernels respectively, next we report the results for the BDS test and the GMR test. The nominal size of the test was set to 0.05. We consider three lags $l = 1, 2, 3$ for delay vector dimension $m = 2$, and only one lag $l = 1$ for higher dimensions ($m = 3, 4, 5, 10$). The bandwidth set H included $K = 5$ different values in the range $R = [0.5, 2.00]$ (after normalization of the series to the unit variance). We used series of length $n = 100$ ($n = 50$ for DGP 6 and $n = 20$ for DGP 8), 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	l	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	l	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	l	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. $y_t = \mathrm{sign}(y_{t-1}) + \varepsilon_t$ | 7. $y_t = 0.6\varepsilon_{t-1}y_{t-2} + \varepsilon_t$

m	l	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. $y_t = 4y_{t-1}(1-y_{t-1}),\ 0 < y_t < 1$					9. $y_t = \varepsilon_t\sqrt{1+0.4y_{t-1}^2}$				
m	l	Q_g	Q_d	Q_c	BDS	GMR	Q_g	Q_d	Q_c	BDS	GMR
2	1	0.98	0.97	1.00	0.85	0.96	0.25	0.22	0.23	0.55	0.32
2	2	0.08	0.12	0.23	0.16	0.74	0.07	0.07	0.06	0.12	0.09
3	1	0.90	0.92	1.00	0.71	0.95	0.24	0.22	0.24	0.51	0.26
5	1	0.54	0.59	0.79	0.35	0.88	0.24	0.20	0.21	0.39	0.18
10	1	0.17	0.24	0.31	0.11	0.59	0.20	0.16	0.18	0.24	0.11

		10. $y_t = \sqrt{h_t}\varepsilon_t,$ $h_t = 0.01 + 0.80h_{t-1} + 0.15y_{t-1}^2$					11. $y_t = \begin{cases} -0.5y_{t-1} + \varepsilon_t, y_{t-1}¡1 \\ 0.4y_{t-1} + \varepsilon_t, \text{else} \end{cases}$				
m	l	Q_g	Q_d	Q_c	BDS	GMR	Q_g	Q_d	Q_c	BDS	GMR
2	1	0.13	0.12	0.12	0.26	0.13	0.91	0.94	0.90	0.07	0.49
2	2	0.12	0.12	0.11	0.23	0.11	0.10	0.09	0.10	0.06	0.05
3	1	0.18	0.15	0.16	0.35	0.15	0.87	0.91	0.87	0.06	0.34
5	1	0.26	0.23	0.24	0.43	0.18	0.77	0.83	0.76	0.04	0.23
10	1	0.38	0.32	0.36	0.48	0.17	0.49	0.61	0.47	0.03	0.14

References

Anderson, N. H., Hall, P. and Titterington, D. M. (1994). Two-Sample Test Statistics for Measuring Discrepancies between Two Multivariate Probability Density Functions Using Kernel-Based Density Estimates. *Journal of Multivariate Analisys* **50**, 41–54.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

Brock, W., Dechert, W., Scheinkman, J. and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews* **15**, 197–236.

Denker, M. and Keller, G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte 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. *Physical Review E* **53**, 2169–2176.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **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. *International Statistical Review* **61**, 419–433.

Good, P. (2000). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Series in Statistics. New York: Springer-Verlag.

Granger, C. and Lin, J. (1994). Using the Mutual Information Coefficient to Identify Lags in Nonlinear Models. *Journal of Time Series Analysis* **15**, 371–384.

Granger, C. W. and Andersen, A. P. (1978). *An Introduction to bilinear time series models*. Göttingen: Vandenhoek and Ruprecht.

Granger, C. W., Maasoumi, E. and Racine, J. (2004). A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* **25**, 649–669.

Granger, C. W. and Teräsvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters* **62**, 161–165.

Hoeffding, W. (1948). A non-parametric test of independence. *The Annals of Mathematical Statistics* **19**, 546–557.

Hoeffding, W. (1952). The Large-Sample Power of Tests Based on Permutations of Observations. *Annals of Mathematical Statistics* **23**, 169–192.

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.

Marriott, F. (1979). Barnard's Monte Carlo tests: How many simulations? *Applied Statistics* **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 Journal of the Royal Statistical Society* **4**, 225–232.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *The Annals of Statistics* **3**, 1–14.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

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. *Journal of Multivariate Analysis* **93**, 58–80.

Tjøstheim, D. (1996). Measures of dependence and tests of independence. *Statistics* **28**, 249–284.

Tong, H. (1978). On a Threshold Model. In *Pattern Recognition and Signal Processing*, C. H. Chen, ed. Amsterdam: Sijhoff and Noordhoff, pp. 101–141.