FUNDAÇÃO GETULIO VARGAS

EPGE

Escola de Pós-Graduação em Economia

Ensaios Econômicos

Escola de Pós-Graduação em Economia da Fundação Getulio Vargas

N° 413 ISSN 0104-8910

Nonparametric Entropy-Based Tests of Independence Between Stochastic Processes

Marcelo Fernandes

Março de 2001

URL: http://hdl.handle.net/10438/957

Fernandes, Marcelo
Nonparametric Entropy-Based Tests of Independence Between Stochastic Processes/ Marcelo Fernandes - Rio de Janeiro : FGV,EPGE, 2010
(Ensaios Econômicos; 413)

Inclui bibliografia.

CDD-330

Nonparametric entropy-based tests of independence between stochastic processes*

Marcelo Fernandes

Graduate School of Economics, Fundação Getulio Vargas

Praia de Botafogo, 190, 22953-900 Rio de Janeiro, Brazil

Tel: +55.21.5595827 Fax: +55.21.5595827

E-mail: mfernand@fgv.br

* This paper revises the third chapter of my Ph.D. dissertation at the Solvay Business School, Université Libre de Bruxelles. I am indebted to Yacine Aït-Sahalia, Christophe Croux, Renato G. Flôres Jr., Han Hong, two anonymous referees, and seminar participants at the Université Libre de Bruxelles, European University Institute and the 4th Spring Meeting of Young Economists (Amsterdam, 1999) for valuable comments. I acknowledge with gratitude the financial support from CNPq-Brazil (grant 200608/95-9) and a Jean Monnet Fellowship at the European University Institute. The usual disclaimer applies.

Entropy-based tests of independence

Marcelo Fernandes

Graduate School of Economics, Fundação Getulio Vargas

Praia de Botafogo, 190, 22953-900 Rio de Janeiro, Brazil

Tel: +55.21.5595827 Fax: +55.21.5595827

E-mail: `mfernand@fgv.br`

Nonparametric entropy-based tests of independence between stochastic processes

Abstract. This paper develops nonparametric tests of independence between two stationary stochastic processes. The testing strategy boils down to gauging the closeness between the joint and the product of the marginal stationary densities. For that purpose, I take advantage of a generalized entropic measure so as to build a class of nonparametric tests of independence. Asymptotic normality and local power are derived using the functional delta method for kernels, whereas finite sample properties are investigated through Monte Carlo simulations.

JEL classification numbers. C12, C14.

Keywords. independence, nonparametric testing, Tsallis entropy.

1 Introduction

Independence is one of the most valuable concepts in econometrics as virtually all tests boil down to checking some sort of independence assumption. Accordingly, there is an extensive literature on how to test independence, e.g. Hoeffding (1948), Baek and Brock (1992), Johnson and McClelland (1998), and Pinkse (1999). Tjøstheim (1996) offers an excellent survey of the literature.

The fact that stochastic processes are potentially path-dependent complicates the task of developing a suitable test. Consider two stochastic processes $\{X_t,\ t \geq 0\}$ and $\{Y_t,\ t \geq 0\}$. The null hypothesis of interest then reads

$$H_0^*: \quad f_{XY}(X_1, X_2, \ldots, Y_1, Y_2, \ldots) = f_X(X_1, X_2, \ldots) f_Y(Y_1, Y_2, \ldots) \ \text{ a.s.}$$

It is infeasible to develop a test without imposing additional structure. For instance, if X_t and Y_t are independent and identically distributed (iid) univariate processes,[1] it then suffices to consider

$$H_0: \quad f_{XY}(X_t, Y_t) = f_X(X_t) f_Y(Y_t) \ \text{ a.s.} \tag{1}$$

Yet, even in the more general setting where $\{X_t,\ t \geq 0\}$ and $\{Y_t,\ t \geq 0\}$ are stationary stochastic processes, the null hypothesis in (1) has an interesting interpretation. As singled out by Phillips (1991), the stationary joint-density f_{XY} corresponds to the stochastic equilibrium of the bivariate processes $\{(X_t, Y_t),\ t \geq 0\}$, hence (1) corresponds to the property of long-run independence (Gregory and Sampson, 1991).

Serial independence is a particular case in which Y_t consists of lagged values of X_t. Robinson (1991) proposes a test based on the closeness of the joint density of (X_t, X_{t-i}) and the product of the marginals of X_t and X_{t-i} as measured by the Kullback-Leibler information. Skaug and Tjøstheim (1993 and 1995) extend Robinson's framework to other measures of discrepancy between densities such as the Hellinger distance. Somewhat related are tests which examine restrictions on the correlation integral (e.g. Baek and Brock, 1992; Brock, Dechert, Scheinkman and LeBaron, 1996; Mizrach, 1995) and on the characteristic functions (Pinkse, 1998). These tests are particularly interesting for diagnostic checking purposes since they are nuisance parameter free (de Lima, 1996; Pinkse, 1998). Rank tests stand as another valuable alternative (Hallin and Puri, 1992; Hallin, Jurečková, Picek and Zahaf, 1997).

This paper proposes tests for independence between two stationary stochastic processes based on (1). The strategy relies on measuring the closeness between kernel estimates of the joint density and the product of the marginal densities. Instead of the conventional Euclidean distance, I employ a generalized entropic measure ρ_q as suggested by Tsallis (1998). This generalized statistic permits to construct a class of nonparametric tests of independence by varying the entropic index q. The motivation is twofold. First, entropy-based tests are quite appealing for having an information-theoretic interpretation. Second, tests based on the Kullback-Leibler information and Hellinger distance, which are particular cases of the Tsallis generalized entropy, seem to compete well in terms of power to tests using quadratic distances (Skaug and Tjøstheim, 1996).

The remainder of this paper is organized as follows. Section 2 describes some useful properties of the generalized Tsallis entropy. Section 3 proposes the class of nonparametric tests of independence I have in mind and provides asymptotic justification. Asymptotic normality is derived using the Aït-Sahalia's (1994) functional delta method both under the null and under a sequence of local alternatives. Further, I demonstrate that the tests are nuisance parameter free, and so suitable to specification testing. Section 4 investigates the finite sample properties of these tests through Monte Carlo simulations. Section 5 discusses briefly how to obtain more accurate critical values (and p-values) through resampling techniques. Section 6 summarizes the main results and offers some concluding remarks. For ease of exposition, an appendix collects technical lemmas and proofs.

2 Generalized entropic measure

In a multifractal framework, Tsallis (1988) generalizes the Boltzmann-Gibbs-Shannon statistics to address non-extensive systems by introducing an entropy which encompasses the standard Kullback-Leibler measure. This generalized entropy reads

$$\rho_q(f,g) \equiv \frac{1}{1-q}\left\{1 - \int \big[g(u)/f(u)\big]^{1-q} f(u)\mathrm{d}u\right\}, \tag{2}$$

where q stands for the entropic index that characterizes the degree of non-extensivity of the system. In the limiting case $q \to 1$, the Tsallis entropy recovers the Kullback-Leibler information

$$\rho_1(f,g) = \int \log\big[f(u)/g(u)\big] f(u)\mathrm{d}u, \tag{3}$$

whereas, for $q = 1/2$, it boils down to

$$\rho_{1/2}(f,g) = \int \left[\sqrt{f(u)} - \sqrt{g(u)}\right]^2 \mathrm{d}u = 2H^2(f,g), \tag{4}$$

where $H(f,g)$ denotes the Hellinger distance between f and g. The latter is known to entail more robustness with respect to contaminated data (e.g. inliers and outliers) than the usual quadratic metric (Pitman, 1979; Hart, 1997).

Varying the entropic index in the Tsallis statistic results in a class of tests for comparing density functionals. Therefore, it is interesting to derive the properties of ρ_q according to the support on which q lies. Tsallis (1998) shows that, if the entropic index is positive, $\rho_q(f,g)$ is non-negative with equality holding if and only if f coincides with g almost everywhere. Moreover,

$$\frac{\rho_q(f,g)}{q} = \frac{\rho_{1-q}(g,f)}{1-q},$$

thus it is enough to consider $q \geq 1/2$. In this range, the Tsallis entropy satisfies three properties that are desirable in a statistic for testing independence, namely, invariance under suitable transformation, i.e. $\rho_q\big[f(u), g(u)\big] = \rho_q\big[f_\ell(v), g_\ell(v)\big]$ for $v = \ell(u)$; $\rho_q(f,g) \geq 0$ (non-negativeness); and $\rho_q(f,g) = 0$ if and only if $f = g$ (consistency). Though it is not a symmetric measure of discrepancy for

$q \neq 1/2$, symmetry is easily achieved by considering the quantity $\rho_q^S(f,g) = \frac{1}{2}\left[\rho_q(f,g) + \rho_q(g,f)\right]$.

3 Asymptotic tests of independence

For ease of exposition, I consider univariate processes $\{X_t,\ t \geq 0\}$ and $\{Y_t,\ t \geq 0\}$ with discretely recorded observations $(X_1, \ldots, X_T)$ and $(Y_1, \ldots, Y_T)$. Although it is straightforward to extend these techniques to consider multivariate processes, it may be not empirically recommendable in view of the 'curse of dimensionality' that plagues nonparametric estimation. I impose the following regularity conditions.

A1 $\{(X_t, Y_t),\ t \in \mathbb{N}\}$ is strictly stationary and β-mixing with $\beta_j = O(j^{-\delta})$ where $\delta > 1$. In addition, the density f_{XY} is such that $E\|(X_t, Y_t)\|^k < \infty$ for some constant $k > 2\delta/(\delta - 1)$.

A2 The density function f_{XY} is continuously differentiable up to the order s and their successive derivatives are bounded and belong to $L^2\left(\mathbb{R}^2\right)$.

A3 The kernel function K is of order s (even integer), and is continuously differentiable up to order s on $\mathbb{R}$ with derivatives in $L^2(\mathbb{R})$.

A4 The bandwidths $b_{x,T}$ and $b_{y,T}$ used in the estimation of f_{XY} are such that $T^{1/2} b_{\cdot,T}^s + T^{-1/2} b_{\cdot,T}^{-m} \to 0$ for every non-negative integer $m < s/2$.

Assumption A1 restricts the amount of dependence allowed in the observed data sequence to ensure that the central limit theorem holds. As usual, it turns out that there is a trade-off between the degree of dependence and the number of finite moments. Assumptions A2 and A3 determine that, in order to use effectively a kernel of order s for bias reduction, the joint density f_{XY} must have at least that many derivatives. Assumption A4 restricts the rate at which the smoothing parameters in the kernel estimation of the joint density f_{XY} must converge to zero.

To test the null hypothesis H_0, I evaluate the generalized entropy ρ_q at the kernel density estimates $\hat{f} = \hat{f}_{XY}$ and $\hat{g} = \hat{f}_X \hat{f}_Y$, namely

$$\hat{\rho}_q = \frac{1}{1-q} \left\{ 1 - \frac{1}{T} \sum_{t=1}^{T} \left[\frac{\hat{f}_X(X_t)\hat{f}_Y(Y_t)}{\hat{f}_{XY}(X_t, Y_t)} \right]^{1-q} \right\}. \tag{5}$$

The corresponding functional reads

$$\Lambda_f = \frac{1}{1-q} \int \left\{ 1 - \left[\frac{g_{XY}(x,y)}{f_{XY}(x,y)} \right]^{1-q} \right\} f_{XY}(x,y)\, \mathrm{d}(x,y), \tag{6}$$

and it follows from the functional delta method that the asymptotic distribution of $\hat{\rho}_q$ is driven by the first non-degenerate functional derivative of Λ_f. It turns out, however, that the first derivative is singular and the limiting distribution implied by the second derivative is well defined only if the stochastic process (X_t, Y_t) takes value in a bounded support, say $\mathcal{S}_{XY}$.

Proposition 1. *Suppose that the bandwidths are of order* $o\left(T^{-1/(2s+1)}\right)$. *Under assumptions A1 to A3, the normalized statistic*

$$\hat{r}_q = \frac{T\, b_{x,T}^{1/2}\, b_{y,T}^{1/2}\, \hat{\rho}_q - b_{x,T}^{1/2}\, b_{y,T}^{1/2}\, \hat{\delta}}{\hat{\sigma}} \xrightarrow{d} N(0,1),$$

where $\hat{\delta}$ *and* $\hat{\sigma}^2$ *are consistent estimators of* $\delta = \left[\int |K(u)|^2\, du\right]^2 \int_{\mathcal{S}_{XY}} d(x,y)$ *and* $\sigma^2 = \left\{\int \left[\int K(u)K(u+v)\, du\right]^2 dv\right\}^2 \int_{\mathcal{S}_{XY}} d(x,y)$, *respectively.*

As is apparent, the asymptotic mean and variance exist only if the support $\mathcal{S}_{XY}$ is bounded.[2] To avoid such a restrictive assumption, it is necessary to contrive some sort of weighting scheme. Consider next the following functional

$$\Lambda_f^w = \frac{1}{1-q} \int w_f(x,y) \left\{ 1 - \left[\frac{g_{XY}(x,y)}{f_{XY}(x,y)} \right]^{1-q} \right\} f_{XY}(x,y)\, \mathrm{d}(x,y), \tag{7}$$

where $w_f(x,y)$ is a general weighting function that may depend on the density $f_{XY}(x,y)$ as in Fan and Li (1996). To establish the limiting distribution of the sample analog of (7), i.e.

$$\hat{\rho}_q^w = \frac{1}{T(1-q)} \sum_{t=1}^{T} w_f(X_t, Y_t) \left\{ 1 - \left[\frac{\hat{f}_X(X_t)\hat{f}_Y(Y_t)}{\hat{f}_{XY}(X_t, Y_t)} \right]^{1-q} \right\}, \tag{8}$$

one additional assumption is necessary.

A5 Consider f^*_{XY} and f^+_{XY} in a neighborhood N_f of the true density f_{XY}. The weighting function $w_f(x, y)$ is separable, i.e. $w_f(x, y) = w_f(x)w_f(y)$, and such that

$$
\begin{aligned}
&(i) \quad \int w_f(x,y)\big(\hat{f}_{XY} - f_{XY}\big)\, \mathrm{d}(x,y) \neq 0, \\
&(ii) \quad E\,|w_f(x,y)|^{3+r} < \infty, \text{ for } r > (3+\epsilon)(3+\epsilon/2)/\epsilon, \ \forall \epsilon > 0, \\
&(iii) \quad E \sup_{f^*_{XY} \in N_f} |w_{f^*}(x,y)|^2 < \infty, \\
&(iv) \quad E\,\big|w_{f^*}(x,y) - w_{f^+}(x,y)\big|^2 \leq c||f^*_{XY} - f^+_{XY}||^2_{L(\infty,m^*)},
\end{aligned}
$$

where c is a constant and m^* is an integer such that $0 < m^* < s/2 + 1/4$.

The first condition of A5 ensures that first functional derivative of Λ^w_f is not degenerate. It excludes, for instance, the trivial case $w_f(x, y) = 1$ considered in Proposition 1. In turn, the other three conditions guarantee that one may truncate the infinite sum that appears in the asymptotic variance of the test statistics. In particular, the trimming function $w(x, y) = \mathbb{1}_{\mathcal{S}}(x, y)$, where $\mathcal{S} = \mathcal{S}^X \times \mathcal{S}^Y$ is a compact subset of the density support, satisfies A5. Lastly, the next result assumes implicitly that Λ^w_f is Fréchet differentiable with respect to the Sobolev norm of order $(2, m)$ at the true joint density.

Before stating the next result, it is useful to establish some notation. Let $\mu_u = E\big[w_f(u_t)\big]$, $\tau_u(k) = E\big[w_f(u_t)w_f(u_{t+k})\big]$, and $\gamma_u(k) = \tau_u(k) - \mu_u^2$. Notice that, under the null of independence, $\mu_{XY} = \mu_X\mu_Y$ and $\tau_{XY}(k) = \tau_X(k)\tau_Y(k)$.

Theorem 1. *Under assumptions A1 to A5, the normalized statistic*

$$\hat{r}^w_q = \frac{\sqrt{T}\,\hat{\rho}^w_q}{\hat{\sigma}_w} \xrightarrow{d} N(0,1),$$

where $\hat{\sigma}^2_w$ is a consistent estimator of the long-run variance-covariance matrix $\sigma^2_w = \sum_{k=-\infty}^{\infty} \big\{\gamma_{XY}(k) + \gamma_X(k)\mu^2_X + \gamma_Y(k)\mu^2_Y - 2\big[\gamma_X(k) + \gamma_Y(k)\big]\mu_{XY}\big\}$.

Ergo, a test which rejects the null hypothesis at the level α when $\hat{r}^w_q$ is greater than or equal to the $(1-\alpha)$-quantile $z_{1-\alpha}$ of a standard normal distribution is locally strictly unbiased. To assess the asymptotic local power, consider a local alternative of the form[3]

$$H_{1,T}: \sup_{(x,y)\in\mathcal{S}} \left| f_{XY}^{[T]}(x,y) - g_{XY}^{[T]}(x,y)[1+(q-1)\epsilon_T \lambda_{XY}(x,y)]^{1/(q-1)} \right|, \tag{9}$$

where $\epsilon_T = T^{-1/2}$ and λ_{XY} is such that $\delta_\lambda = E\big[w_f(x,y)\lambda_{XY}(x,y)\big]$ exists.

Proposition 2. *Under assumptions A1 to A5, the asymptotic local power is given by* $\Pr\left(\hat{r}_q^w \geq z_{1-\alpha} \middle| H_{1,T}\right) \longrightarrow 1 - \Phi\left(z_{1-\alpha} - \delta_\lambda/\sigma_w\right)$.

Unfortunately, the asymptotic local powers obtained by tests based on different entropic indexes q cannot be directly compared since the local alternatives become closer to the null as q increases.

How to select the weighting scheme is an arbitrary task. Previous works which deal with entropy-based tests of serial independence use simple weighting schemes to preserve the information-theoretic interpretation. For instance, Skaug and Tjøstheim (1996) show that tests based on the Hellinger distance and the Kullback-Leibler information compete well in power against tests based on quadratic measures even for a simple trimming function that bounds the observations to some compact set $\mathcal{S} = \mathcal{S}^X \times \mathcal{S}^Y$ strictly contained in the support of the density. In turn, Robinson (1991) and Pinkse (1994) adopt the following sample-splitting weighting scheme

$$w_t(x,y) = \begin{cases} \mathbb{1}_{\mathcal{S}}(x,y)(1+\gamma) & \text{if } t \text{ is odd} \\ \mathbb{1}_{\mathcal{S}}(x,y)(1-\gamma) & \text{if } t \text{ is even.} \end{cases} \tag{10}$$

As the latter design seems to produce tests with low power against both fixed (Drost and Werker, 1993) and local alternatives (Hong and White, 2000), I follow Skaug and Tjøstheim's simpler approach that relies on a separable trimming function.

3.1 Serial independence

Testing for serial independence stands for an interesting application of tests of independence. Consider, for instance, a process $\{X_t;\ t \in \mathbb{N}\}$. Serial independence implies that the joint distribution of the realizations of the process coincides almost everywhere with the product of the marginal distributions, i.e.

$$\Pr(X_0, \ldots, X_t) = \Pr(X_0) \ldots \Pr(X_t) \;\text{ a.s.} \tag{11}$$

For the sake of feasibility, it is convenient to work with a pairwise approach, i.e. to test independence between pairs, say (X_t, X_{t-i}). Thus, the resulting null hypothesis is only a necessary condition for serial independence, namely

$$H_0^i : f(X_t, X_{t-i}) = f(X_t) f(X_{t-i}) \;\text{ a.s.}, \tag{12}$$

where $f(X_t, X_{t-i})$, $f(X_t)$ and $f(X_{t-i})$ denote the joint density of (X_t, X_{t-i}), and the marginal densities of X_t and X_{t-i}, respectively.

It follows immediately from Theorem 1 that a test which rejects the null hypothesis H_0^i at the level α when $\sqrt{T}\,\hat{\rho}^w_{q,i} \geq z_{1-\alpha}\,\hat{\gamma}_X(0)$, where $\hat{\gamma}_X(0)$ is a consistent estimator of $\gamma_X(0) = \mathrm{Var}\big[w_f(X_t)\big]$ and

$$\hat{\rho}^i_{q,w} = \frac{1}{(1-q)(T-i)} \sum_{t=i+1}^{T} w_f(X_t, X_{t-i}) \left\{ 1 - \left[\frac{\hat{f}(X_t)\hat{f}(X_{t-i})}{\hat{f}(X_t, X_{t-i})} \right]^{1-q} \right\} \tag{13}$$

is locally strictly unbiased.

Corollary. *Under assumptions A1 to A5, the normalized statistic*

$$\hat{r}^w_{q,i} = \frac{\sqrt{T}\,\hat{\rho}^w_{q,i}}{\hat{\gamma}_X(0)} \xrightarrow{d} N(0,1),$$

where $\hat{\gamma}_X(0)$ *is a consistent estimator of* $\gamma_X(0) = Var\big[w_f(X_t)\big]$.

Failing to reject H_0^1 indicates that X_t does not depend significantly on X_{t-1}, but it could well depend on another past realization, say X_{t-4}. The simplicity of the pairwise approach comes at the expense of an uncomfortable dependence on lags. Yet, one can mitigate this dependence by considering a null hypothesis such as $H_0^s : \cap_{n=1}^N H_0^{i_n}$ $(i_1 < \ldots < i_N)$ as in Skaug and Tjøstheim (1996). In particular, it is possible to demonstrate that the sum statistic $\hat{\rho}^s_{q,w} = \sqrt{T}\,\sum_{n=1}^N \hat{\rho}^w_{q,i_n}$ is asymptotically normal with mean zero and variance $N\gamma_X^2(0)$.

3.2 Specification testing and nuisance parameters

It is often the case that the process of interest is unobservable. In specification testing, for instance, one usually examines whether the residuals are iid. Serial dependence may indicate that a lagged dependent variable was omitted, whereas if homoskedasticity does not hold, one may wish to model the form of heteroskedasticity to increase the efficiency of the estimation. Suppose that there exists an observable vector series $(X_1, \ldots, X_T)$ and a function ξ known up to a parameter vector θ such that $Y_t = Y_t(\theta) = \xi(X_t, \theta)$, $t = 1, \ldots, T$. In this setting, the interest is in testing model specification by checking whether the error term $Y_t = Y_t(\theta)$ is serially independent. Of course, feasible testing procedures rely on a consistent estimate $\hat{\theta}$ of the parameter vector θ so as to form the series of residuals $\hat{Y}_t = Y_t\big(\hat{\theta}\big)$, $t = 1, \ldots, T$.

The next result establishes the conditions in which the entropy-based tests of independence are nuisance parameter free and hence there is no asymptotic cost in substituting residuals for errors. It turns out that the requirements are very mild.

Theorem 2. *Under assumptions A1 to A5, the normalized statistic*

$$\hat{r}^w_{q,i}\big(\hat{\theta}\big) = \frac{\sqrt{T}\,\hat{\rho}^w_{q,i}\big(\hat{\theta}\big)}{\hat{\gamma}_{\hat{Y}}(0)} \xrightarrow{d} N(0,1),$$

where $\hat{\theta}$ is a T^d-consistent estimator of θ with $d \geq \max\left\{\frac{2}{s+1} - \frac{1}{2}, \frac{3}{2(s+1)} - \frac{1}{4}\right\}$.

The condition on the rate of convergence gets more stringent as the order s of the kernel decreases, conforming with the fact that higher-order kernels converge at a faster rate. Accordingly, it suffices to verify that the condition reduces to $d \geq 1/4$ for second-order kernels to conclude that little is required in Theorem 2. Indeed, it is difficult to think of any reasonable estimator that does not satisfy such condition.

4 Finite sample properties

There are two prime reasons to believe that the asymptotic theory of entropy-based tests performs poorly in finite samples. First, the error of neglecting higher-order terms may be substantial in the event that these terms are close in order to the dominant term (Fan and Linton, 1997; Skaug and Tjøstheim, 1993). Second, for the particular case in which the weighting function simply trims data out of a compact set, boundary effects may disrupt the asymptotic approximation. As the support grows, the variance of the limiting distribution increases, whereas the estimate of the test statistic remains unaltered once all of the observations are included. Therefore, it will be not surprising if asymptotic tests turn out to work unsatisfactorily in small samples.

In what follows, I perform a limited Monte Carlo exercise to assess the performance of entropy-based tests in finite samples. All results are based on 2000 replications and consider two sample sizes. To avoid initialization problems, I simulate 1500 realizations of each data generating process and take the last 500 and 1000 observations to compute the test statistics with entropic index $q \in \{1/2, 1, 2, 4\}$. For simplicity, I utilize a trimming function $w(x,y) = \mathbb{1}_{\mathcal{S}}(x,y)$ that allocates weight zero to observations out of the compact set $\mathcal{S} = \mathcal{S}^X \times \mathcal{S}^Y$, where $\mathcal{S}^u = \big\{u : \, \big|u - \bar{u}\big| < 2\hat{s}_u\big\}$ with $\bar{u}$ and $\hat{s}_u^2$ denoting the sample mean and variance, respectively. Further, all kernel density estimations are carried out using a Gaussian kernel and the bandwidth recommended by Silverman's (1986) rule of thumb.

To examine the size properties of the entropy-based tests, I rely on a simple specification where X_t and Y_t are independent Gaussian autoregressive processes of order one, AR(1). More precisely, the data generating mechanism reads

$$
\begin{aligned}
Y_t &= 0.8\, Y_{t-1} + \upsilon_t, \qquad \upsilon_t \sim \text{ iid } N(0,1) \\
X_t &= 0.8\, X_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0,1),
\end{aligned}
$$

where ϵ_t and υ_s are independent for every t and s. The results in Table 1 indicates that the critical values given by the asymptotic approximation are of little value. The reasons are twofold. First, since ρ_q is non-negative, it turns out that it is seldom the case that the normalized test statistics take negative values. In fact, the degree of non-normality seems to increase with the entropic index q, suggesting that lower entropic indexes entail more robust test statistics. Second, the variances of the test statistics, which are computed using the Newey and West's (1987) estimator with Andrews's (1991) automatic bandwidth, are systematically overestimated. Further simulations point out that this pattern is quite robust to variations in the bandwidth as opposed to variations in the autoregressive coefficient. As expected, the performance of the asymptotic approximation improves as one reduces the data persistence.

Table 2 and 3 document the finite sample power of the entropy-based tests against alternatives characterized by dependence in mean and in variance, respectively. The former is represented by letting Y_t follow an autoregressive distributed lag ADL(1,0) process, namely

$$\begin{aligned} Y_t &= 0.8\,Y_{t-1} + X_t + \upsilon_t, & \upsilon_t &\sim \text{ iid } N(0,1) \\ X_t &= 0.8\,X_{t-1} + \epsilon_t, & \epsilon_t &\sim \text{ iid } N(0,1). \end{aligned}$$

For the alternative that imposes dependence in variance, I utilize an autoregressive process with heteroskedastic error, viz.

$$\begin{aligned} Y_t &= 0.8\,Y_{t-1} + X_t\,\upsilon_t, & \upsilon_t &\sim \text{ iid } N(0,1) \\ X_t &= 0.8\,X_{t-1} + \epsilon_t, & \epsilon_t &\sim \text{ iid } N(0,1). \end{aligned}$$

The figures concerning the (size-corrected) power are quite rosy, confirming the competitiveness of the entropy-based tests of independence. The snag is that one does not know the proper critical values in finite samples and the asymptotic approximation performs very poorly. It is therefore paramount to contrive a procedure that engenders more accurate critical values for the tests. I defer this issue to Section 5, where I discuss suitable resampling techniques.

4.1 Serial independence

Next I move to investigating whether the asymptotic tests of serial independence (with lag $i = 1$) have the correct size. I generate random variables coming from a standard normal, a chi-squared distribution with one degree of freedom, and a t-student with 5 degrees of freedom. The second distribution exhibits highly positive skewness whereas the third is known to display thick tails, i.e. leptokurtosis. To inspect how powerful these tests are in finite samples, I rely on two simple data generating mechanisms, namely an AR(1) and an autoregressive conditional heteroskedastic model of order one, ARCH(1). The former deals with serial dependence in the mean and evolves according to $X_t = 0.8\, X_{t-1} + \epsilon_t$. In contrast, the ARCH(1) explores the case in which there is no serial correlation, though the process exhibits serial dependence in the second moment. More precisely, it follows $X_t = \left(0.2 + 0.8\, X_{t-1}^2\right)^{1/2} \epsilon_t$, where the error ϵ_t has a standard normal distribution given the past realizations of X_t. The size-corrected power of the entropy-based tests against AR(1) and ARCH(1) processes are easily computed using the critical values in Table 5.

Tables 4 to 6 report some descriptive statistics concerning the distribution of the normalized test statistics when the null hypothesis is true. For the standard normal iid case in Table 4, the distributions are roughly normal, for all entropic indexes, i.e. skewness and kurtosis are not far from zero and three, respectively. However, there is a poor correspondence between the asymptotic mean and variance of the test statistics and their simulated counterparts. Similar patterns also emerge in non-normal iid random variables (see Tables 5 and 6). If, on the one hand, it conforms with the results of Skaug and Tjøstheim (1996); on the other hand, the Gaussian character of the finite sample distributions is sort of surprising for, in general, smoothing-based tests resemble more closely parametric chi-squared tests. Indeed, Staniswalis and Severini (1991) and Hjellvik and Tjøstheim (1996b), among others, propose the use of chi-squared and gamma approximations to cope with the large bias and skewness that are typically revealed by Monte Carlo experiments (Fan, 1995; Hjellvik and Tjøstheim, 1996a).

Tables 7 and 8 documents the size-corrected power of the nonparametric entropy-based tests in finite samples. For the autoregressive process, the distributions are fairly normal for all entropic indexes, whereas the distributions are farther from normality in the ARCH(1) case. Close inspection reveals however that there are two outliers in the latter that makes the even moments take extremely high values. At any rate, the size-corrected power of the tests are excellent for both alternatives irrespective of the entropic index.

5 Resampling methods

The finite sample analysis in the previous section singles out that the asymptotic critical values of the entropy-based tests are not reliable. Moreover, additional simulations reveal that the finite sample distributions of the test statistics depend heavily on the bandwidth of the kernel density estimation as in Skaug and Tjøstheim (1993). Therefore, in what follows, I discuss some refinements in the testing procedures in order to ameliorate the accuracy of the critical values.

Under the independence between X and Y, it seems natural to apply bootstrap techniques to compute appropriate critical values. In principle, one simply needs to resample from the empirical marginal distributions of X and Y thereby imposing independence. However, X and Y are weakly dependent stationary time series, and thus one must employ a resampling scheme suitable to dependent data. As the testing procedure relies on kernel density estimation, it seems convenient to use Politis and Romano's (1994) stationary bootstrap to ensure the stationarity of the bootstrap samples. Politis and Romano establish its asymptotic validity under the assumption that the original statistic is asymptotically normal under the null and the absence of nuisance parameters, whereas White (1999) extends the result to statistics with nuisance parameters.

The usual block bootstrap procedure (Hall, 1985; Künsch, 1989; Liu and Singh, 1992) provides artificial time series which are not stationary due to the difference in the joint distribution of resampled observations close to a join between blocks and observations in the centre of a block. Similar to block resampling schemes, the stationary bootstrap resamples by blocks the original data in order to form pseudo-time series from which the test statistic may be recalculated. However, instead of fixing the size of the blocks, the stationary bootstrap takes blocks of random length m. More specifically, Politis and Romano suggest the use of a geometric distribution

$$\Pr(m = j) = (1 - p)^{j-1} p, \quad j = 1, 2, \ldots$$

in order to produce artificial time series which are stationary with mean block length $\ell = 1/p$.[4]

The choice of p is a smoothing issue which has not been theoretically solved. On the one hand, the blocks should be long enough to capture as faithfully as possible the original time dependence of the series. On the other hand, the number of bootstrap samples should be large enough to provide a good estimate of the test statistic distribution, and this points towards short blocks. The few theoretical results available in the literature indicates that a good compromise is achieved by taking p_T of order $T^{-\zeta}$ for some ζ in the interval $(0, 1)$. In addition, restricting ζ to the interval $(0, 1/2)$ suffices to ensure tightness of the bootstrap empirical process (Politis and Romano, 1994).

To assess the performance of the stationary bootstrap, I revisit the first experiment of the previous section where X_t and Y_t follow independent Gaussian AR(1) processes. The processes are equally persistent with autoregressive coefficient ϕ varying from 0.4 to 0.95. To conserve on computation time, the number of replications and bootstrap samples are set to 500 and 99, respectively. The mean block length ℓ of the stationary bootstrap is chosen according to Carlstein's (1986) rule of thumb, namely

$$\ell = \left(\frac{\sqrt{6}\,\hat{\phi}}{1-\hat{\phi}^2}\right)^{-2/3} T^{-1/3},$$

where $\hat{\phi}$ denotes the first-order sample autocorrelation. As before, I consider two sample sizes, namely 500 and 1000 observations.

Table 9 displays quite encouraging results. Despite the reduced number of artificial samples ($B = 99$), the stationary bootstrap mitigates significantly the size distortions especially when data are not very persistent. More precisely, at the 5% level, bootstrap-based tests with low entropic indexes have an empirical size varying from 6% to 16% according to the degree of persistence. As expected, the ability of the stationary bootstrap to mimic the data dependence deteriorates as the persistence increases.

Under the null hypothesis of serial independence, one can also rely on the fact that the order statistic $X^{(\cdot)} = \left(X^{(1)}, \ldots, X^{(T)}\right)$ is a sufficient statistic to justify the use of permutation tests. It is well known that the conditional distribution of $(X_1, \ldots, X_T)$ given $x^{(\cdot)} = \left(x^{(1)}, \ldots, x^{(T)}\right)$ is discretely uniform over the T permutations of $x^{(1)}, \ldots, x^{(T)}$. Then, the conditional distribution of the test statistic $\hat{\rho}_q$ is constructed by evaluating $\hat{\rho}_q$ at each of these T permutations. The critical value

$$c_\alpha^{X^{(\cdot)}} = \inf_c \left\{c : \Pr\left(\hat{\rho}_q \geq c \,\middle|\, X^{(\cdot)}\right) \leq \alpha\right\} \tag{14}$$

provides a permutation test with exact level α. In practice, however, it is impossible to compute exact critical values unless the sample size T is very small. Notwithstanding, an approximation can be obtained by Monte Carlo simulations without any effect on the level of the test — of course, the same cannot be said about the power of the test, which unfortunately decreases.

6 Summary and conclusions

This paper develops a family of nonparametric entropy-based tests of independence in a strictly stationary time-series context. The tests hinge on a class of discrepancy measures implied by the Tsallis generalized entropy to gauge the distance between density functionals. In particular, the asymptotic theory I derive in Section 3 extends in a number of ways Robinson's (1991) and Skaug and Tjøstheim's (1996) results for entropy-based tests of serial independence.

In discussing the advantages and drawbacks of these testing procedures, three remarks are in place. First, the fact that these tests are nuisance parameter free indicates that they might be useful to check model specification. Second, the numerical results reported in Section 4 suggest that the asymptotic approximation performs very poorly in finite samples, which points towards the use of resampling techniques as to mitigate size distortions. Albeit the stationary bootstrap seems to perform reasonably well when both stochastic processes follow a simple autoregressive process, further research is necessary to verify whether that remains valid for more complex data generating mechanism. Third, it is not clear how to select the entropic index q so as to maximize the power of the tests, though statistics with lower entropic indexes appear to engender more powerful and robust tests. Notwithstanding, Tsallis's (1998) conjecture that the optimal entropic index varies in function of the data complexity still needs to be confirmed.

Appendix

Lemma 1. Suppose that the functional

$$\Lambda_f = \frac{1}{1-q} \int \left[1 - \left(g_{XY}(x,y)/f_{XY}(x,y)\right)^{1-q}\right] f_{XY}(x,y)\,\mathrm{d}(x,y)$$

is twice Fréchet-differentiable relative to the Sobolev norm of order $(2,m)$ at the true density function f. Then, under the null hypothesis, the following expansion holds

$$\Lambda_{\hat{f}} = \frac{q}{2} \int \left(\frac{h_{x,y}^2}{f_{x,y}} - \frac{h_x^2}{f_x} - \frac{h_y^2}{f_y} \right) \mathrm{d}(x,y) + O\left(\|h_{x,y}\|_{(2,m)}^3 \right),$$

where $f_{x,y} = f_{XY}(x,y)$, $f_x = f_X(x)$, $f_y = f_Y(y)$, $h_{x,y} = \hat{f}_{XY}(x,y) - f_{XY}(x,y)$, $h_x = \int h_{x,y}\,\mathrm{d}y = \hat{f}_X(x) - f_X(x)$, and $h_y = \int h_{x,y}\,\mathrm{d}x = \hat{f}_Y(y) - f_Y(y)$.

Proof. Let $f_{x,y} = f_{XY}(x,y)$, $f_x = f_X(x)$ and $f_y = f_Y(y)$. By assumption, the functional

$$\Lambda_f = \frac{1}{1-q} \int \left[1 - \left(g_{x,y}/f_{x,y}\right)^{1-q}\right] f_{x,y}\,\mathrm{d}(x,y)$$

admits a second order Taylor expansion, i.e.

$$\Lambda_{f+h} = 7\Lambda_f + \mathrm{D}\Lambda_f(h) + \frac{1}{2}\mathrm{D}^2\Lambda_f(h,h) + O\left(\|h\|_{(2,m)}^3 \right),$$

where $\|\cdot\|_{(2,m)}$ denotes the Sobolev norm of order $(2,m)$. Under the null, it turns out that both Λ_f and its first derivative equal zero. To appreciate the latter, recall that Fréchet differentials can be computed as Gâteaux differentials, i.e. $\mathrm{D}\Lambda_f(h) = \frac{\partial}{\partial\lambda}\Lambda_{f,h}(0)$, where

$$\Lambda_{f,h}(\lambda) = \frac{1}{1-q} \int \left[1 - \left(\frac{g_{x,y}(\lambda)}{f_{x,y} + \lambda h_{x,y}} \right)^{1-q} \right] (f_{x,y} + \lambda h_{x,y})\,\mathrm{d}(x,y)$$

and

$$g_{x,y}(\lambda) = \int \left(f_{x,y} + \lambda h_{x,y}\right) \mathrm{d}y \int \left(f_{x,y} + \lambda h_{x,y}\right) \mathrm{d}x = \left(f_x + \lambda h_x\right)\left(f_y + \lambda h_y\right).$$

Taking the Gâteaux derivative of $g_{x,y}(\lambda)$ yields

$$\begin{aligned} \frac{\partial g_{x,y}(\lambda)}{\partial\lambda} &= h_x \int (f_{x,y} + \lambda h_{x,y})\,\mathrm{d}x + h_y \int (f_{x,y} + \lambda h_{x,y})\,\mathrm{d}y \\ \frac{\partial g_{x,y}(0)}{\partial\lambda} &= f_x h_y + f_y h_x, \end{aligned}$$

which means that the Fréchet derivative of $g_{x,y}$ evaluated at $f_{x,y}$ is simply $\mathrm{D}g_{x,y} = f_x h_y + f_y h_x$. Similarly,

$$
\begin{aligned}
\frac{\partial \Lambda_{f,h}(\lambda)}{\partial \lambda} &= \frac{q}{q-1} \int \left[\frac{g_{x,y}(\lambda)}{f_{x,y} + \lambda h_{x,y}}\right]^{1-q} h_{x,y}\, \mathrm{d}(x,y) \\
&\quad - \int \frac{\partial g_{x,y}(\lambda)}{\partial \lambda} \left[\frac{f_{x,y} + \lambda h_{x,y}}{g_{x,y}(\lambda)}\right]^{q} \mathrm{d}(x,y), \\
\frac{\partial \Lambda_{f,h}(0)}{\partial \lambda} &= \frac{q}{q-1} \int \left(\frac{f_{x,y}}{g_{x,y}}\right)^{q-1} h_{x,y}\, \mathrm{d}(x,y) - \int \mathrm{D}g_{x,y} \left(\frac{f_{x,y}}{g_{x,y}}\right)^{q} \mathrm{d}(x,y).
\end{aligned}
$$

Under the null, the first functional derivative reads

$$
\begin{aligned}
\mathrm{D}\Lambda_f(h) &= \frac{q}{q-1} \int h_{x,y}\, \mathrm{d}(x,y) - \int \mathrm{D}g_{x,y}\, \mathrm{d}(x,y) \\
&= -\int (f_x h_y + f_y h_x)\, \mathrm{d}(x,y) \\
&= -\int f_x h_y\, \mathrm{d}(x,y) - \int f_y h_x\, \mathrm{d}(x,y) \\
&= -\int f_x\, \mathrm{d}x \int h_y\, \mathrm{d}y - \int f_y\, \mathrm{d}y \int h_x\, \mathrm{d}x \;=\; 0,
\end{aligned}
$$

as $\int h_{x,y}\, \mathrm{d}(x,y) = \int h_x\, \mathrm{d}x = \int h_y\, \mathrm{d}y = 0$. It remains to compute the second functional derivative. Note that

$$
\frac{\partial^2 g_{x,y}(\lambda)}{\partial \lambda^2} = 2 h_x h_y,
$$

which implies that $\mathrm{D}^2 g_{x,y} = 2 h_x h_y$. Accordingly,

$$
\begin{aligned}
\frac{\partial^2 \Lambda_{f,h}(\lambda)}{\partial \lambda^2} &= q \int g_{x,y}^{1-q}(\lambda)(f_{x,y} + \lambda h_{x,y})^{q-2} h_{x,y}^2\, \mathrm{d}(x,y) \\
&\quad - q \int g_{x,y}^{-q}(\lambda) \frac{\partial g_{x,y}(\lambda)}{\partial \lambda} (f_{x,y} + \lambda h_{x,y})^{q-1} h_{x,y}\, \mathrm{d}(x,y) \\
&\quad - \int \mathrm{D}^2 g_{x,y} \left[\frac{f_{x,y} + \lambda h_{x,y}}{g_{x,y}(\lambda)}\right]^{q} \mathrm{d}(x,y) \\
&\quad - q \int g_{x,y}^{-q}(\lambda) \frac{\partial g_{x,y}(\lambda)}{\partial \lambda} (f_{x,y} + \lambda h_{x,y})^{q-1} h_{x,y}\, \mathrm{d}(x,y) \\
&\quad + q \int g_{x,y}^{-q-1}(\lambda) \left[\frac{\partial g_{x,y}(\lambda)}{\partial \lambda}\right]^{2} (f_{x,y} + \lambda h_{x,y})^{q}\, \mathrm{d}(x,y), \\
\frac{\partial^2 \Lambda_{f,h}(0)}{\partial \lambda^2} &= \int \left(\frac{g_{x,y}}{f_{x,y}}\right)^{1-q} \frac{h_{x,y}^2}{f_{x,y}}\, \mathrm{d}(x,y) \\
&\quad - 2q \int \mathrm{D}g_{x,y} \left(\frac{f_{x,y}}{g_{x,y}}\right)^{q} \frac{h_{x,y}}{f_{x,y}}\, \mathrm{d}(x,y) \\
&\quad - \int \mathrm{D}^2 g_{x,y} \left(\frac{f_{x,y}}{g_{x,y}}\right)^{q} \mathrm{d}(x,y)
\end{aligned}
$$

$$+ q \int \frac{\left[\mathrm{D} g_{x,y}\right]^2}{f_{x,y}} \left(\frac{f_{x,y}}{g_{x,y}}\right)^{1+q} \mathrm{d}(x,y).$$

Under the null hypothesis, it yields

$$\begin{aligned}
\mathrm{D}^2 \Lambda_f(h,h) &= q \int \frac{h_{x,y}^2}{f_{x,y}} \mathrm{d}(x,y) - 2q \int \mathrm{D} g_{x,y} \frac{h_{x,y}}{f_{x,y}} \mathrm{d}(x,y) \\
&\quad - \int \mathrm{D}^2 g_{x,y} \, \mathrm{d}(x,y) + q \int \frac{\left[\mathrm{D} g_{x,y}\right]^2}{f_{x,y}} \mathrm{d}(x,y) \\
&= q \int \frac{h_{x,y}^2}{f_{x,y}} \mathrm{d}(x,y) - 2q \int \frac{(f_x h_y + f_y h_x) h_{x,y}}{f_{x,y}} \mathrm{d}(x,y) \\
&\quad - 2 \int h_x h_y \, \mathrm{d}(x,y) + q \int_{x,y} \frac{\left(f_x h_y + f_y h_x\right)^2}{f_{x,y}} \mathrm{d}(x,y) \\
&= q \int \frac{h_{x,y}^2}{f_{x,y}} \mathrm{d}(x,y) - 2q \int \frac{f_x h_y h_{x,y}}{f_{x,y}} \mathrm{d}(x,y) \\
&\quad - 2q \int \frac{f_y h_x h_{x,y}}{f_{x,y}} \mathrm{d}(x,y) - 2 \int h_x \, \mathrm{d}x \int h_y \, \mathrm{d}y \\
&\quad + q \int \frac{f_x^2 h_y^2 + 2 f_x f_y h_x h_y + f_y^2 h_x^2}{f_{x,y}} \mathrm{d}(x,y) \\
&= q \int \frac{h_{x,y}^2}{f_{x,y}} \mathrm{d}(x,y) - 2q \int \frac{h_{x,y} h_y}{f_y} \mathrm{d}(x,y) \\
&\quad - 2q \int \frac{h_{x,y} h_x}{f_x} \mathrm{d}(x,y) + q \int \frac{f_x h_y^2}{f_{x,y}} \mathrm{d}(x,y) \\
&\quad + 2q \int \frac{f_x f_y h_x h_y}{f_{x,y}} \mathrm{d}(x,y) + q \int \frac{f_y^2 h_x^2}{f_{x,y}} \mathrm{d}(x,y) \\
&= q \int \frac{h_{x,y}^2}{f_{x,y}} \mathrm{d}(x,y) - 2q \int \frac{h_y^2}{f_y} \mathrm{d}y - 2q \int \frac{h_x^2}{f_x} \mathrm{d}x \\
&\quad + q \int \frac{h_y^2}{f_y} \mathrm{d}y + 2q \int h_x \, \mathrm{d}x \int h_y \, \mathrm{d}y + q \int \frac{f_y h_x^2}{f_x} \mathrm{d}(x,y) \\
&= q \int \left(\frac{h_{x,y}^2}{f_{x,y}} - \frac{h_x^2}{f_x} - \frac{h_y^2}{f_y} \right) \mathrm{d}(x,y),
\end{aligned}$$

which completes the proof.

Proof of Proposition 1. It follows from Aït-Sahalia's functional delta method that, as long as the remainder term in Lemma 1 is bounded after proper normalization, the asymptotic distribution of $T \, b_{x,T}^{1/2} \, b_{y,T}^{1/2} \, \Lambda_{\hat{f}}$ is driven by the second-order functional derivative. For simplicity, suppose that $b_T = b_{x,T} = b_{y,T}$. It is straightforward to show that

$$\left\| \hat{f}_{x,y} - f_{x,y} \right\| = O_p\left(b_T^s + T^{-1/2} b_T^{-1} \log T \right)$$

under assumptions A1 to A3 given the order of the bandwidth (Bosq, 1996). Further, the bandwidth condition ensures the boundness of the remainder term, viz.

$$Tb_T \left\| \hat{f}_{x,y} - f_{x,y} \right\|^3 = O_p\left(Tb_T^{3s+1} + T^{-1/2} b_T^{-2} \log^3 T\right) = o_p(1).$$

Due to the different rates of convergence, it is clear that $\int h_{x,y}^2 / f_{x,y} \, \mathrm{d}(x,y)$ is the leading term in the second functional derivative. To derive the limiting distribution, one may apply the second-order asymptotic theory provided by Aït-Sahalia (1994), which considers Khashimov's (1992) generalization of Hall's (1984) central limit theorem for degenerate U-statistics to weak dependent processes. More precisely, under assumptions A1 to A3 and the bandwidth condition, it follows that

$$T\, b_T \left(U_T - \frac{1}{b_T} \left[\int K^2(u)\, \mathrm{d}u \right]^2 \int \varphi(x,y) f_{XY}(x,y)\, \mathrm{d}(x,y) \right) \xrightarrow{d} N\big(0, V_U\big),$$

where $U_T = \int \varphi(x,y) \left[\hat{f}_{XY}(x,y) - f_{XY}(x,y) \right]^2 \mathrm{d}(x,y)$ and

$$V_U = 2 \left\{ \int \left[\int K(u) K(u+v) \mathrm{d}u \right]^2 \mathrm{d}v \right\}^2 \int \big[\varphi(x,y) f_{XY}(x,y) \big]^2 \mathrm{d}(x,y).$$

As $\varphi(x,y) = f_{XY}^{-1}(x,y)$ in the case under study, a well-defined limiting distribution exists only if the support of f_{XY} is bounded.

Lemma 2. Under the null and assumption A5(i), the following expansion holds

$$\Lambda_{\hat{f}} = \int w_f(x,y) \big(h_{x,y} - f_x h_y - f_y h_x \big)\, \mathrm{d}(x,y) + O\left(\|h_{x,y}\|^2_{(2,m)} \right).$$

Proof: Expanding the functional Λ_f^w yields

$$\begin{aligned} \Lambda_{f,h}^w(\lambda) &= \frac{1}{1-q} \int w_{f+\lambda h} \left[(f+\lambda h)^{1-q} - g_\lambda^{1-q} \right] (f+\lambda h)^q \, \mathrm{d}(x,y) \\ &= \frac{1}{1-q} \int w_{f+\lambda h} \left[f + \lambda h - g_\lambda^{1-q} (f+\lambda h)^q \right] \mathrm{d}(x,y), \end{aligned}$$

where $g_\lambda = g_{x,y}(\lambda)$. It then follows that

$$\frac{\partial \Lambda_{f,h}^w(\lambda)}{\partial \lambda} = \frac{1}{1-q} \int \frac{\partial w_{f+\lambda h}}{\partial \lambda} \left[f + \lambda h - g_\lambda^{1-q} (f+\lambda h)^q \right] \mathrm{d}(x,y)$$

$$
\begin{aligned}
& + \frac{1}{1-q} \int w_{f+\lambda h}\, h_{x,y}\, \mathrm{d}(x,y) \\
& - \int w_{f+\lambda h} \frac{\partial g_\lambda}{\partial \lambda} \left(\frac{f+\lambda h}{g_\lambda} \right)^q \mathrm{d}(x,y) \\
& - \frac{q}{1-q} \int w_{f+\lambda h} \left(\frac{g_\lambda}{f+\lambda h} \right)^{1-q} h_{x,y}\, \mathrm{d}(x,y), \\
\frac{\partial \Lambda^w_{f,h}(0)}{\partial \lambda} \;=\; & \frac{1}{1-q} \int \mathrm{D} w_f(x,y) \big[f_{x,y} - g^{1-q}_{x,y} f^q_{x,y} \big]\, \mathrm{d}(x,y) \\
& + \frac{1}{1-q} \int w_f(x,y) h_{x,y}\, \mathrm{d}(x,y) \\
& - \int w_f(x,y) \mathrm{D} g_{x,y} \left(\frac{f_{x,y}}{g_{x,y}} \right)^q \mathrm{d}(x,y) \\
& - \frac{q}{1-q} \int w_f(x,y) \left(\frac{g_{x,y}}{f_{x,y}} \right)^{1-q} h_{x,y}\, \mathrm{d}(x,y),
\end{aligned}
$$

where $\mathrm{D}w_f$ denotes the functional derivative of $w_f(x,y)$ evaluated at $f_{x,y}$. As the first term is zero under the null hypothesis, the first functional derivative reads

$$
\mathrm{D}\Lambda^w_f(h) = \int w_f(x,y) \big(h_{x,y} - \mathrm{D} g_{x,y} \big)\, \mathrm{d}(x,y).
$$

The result then ensues from assumption A5(i) and the fact that $\mathrm{D}g_{x,y} = f_x h_y + f_y h_x$.

Proof of Theorem 1. Define the vector process $\{A_t,\ t \geq 0\}$, where

$$
A'_t = \{ w_f(X_t, Y_t) - \mu_{XY}, w_f(X_t) - \mu_X, w_f(Y_t) - \mu_Y \}.
$$

By assumption A1, $\{A_t\}$ is also β-mixing and therefore it follows from the central limit theorem for β-mixing processes (Aït-Sahalia, 1994, Lemma 1) that $T^{-1/2} \sum_{t=1}^T A_t \xrightarrow{d} N(0, \Omega)$, where $\Omega = \sum_{k=-\infty}^{\infty} E\big[A_t A'_{t+k}\big]$. It is straightforward to verify that, under the null of independence,

$$
E\big[A_t A'_{t+k}\big] = \begin{pmatrix} \gamma_{XY}(k) & \gamma_X(k)\mu_Y & \gamma_Y(k)\mu_X \\ \gamma_X(k)\mu_Y & \gamma_X(k) & 0 \\ \gamma_Y(k)\mu_X & 0 & \gamma_Y(k) \end{pmatrix}.
$$

Using the expansion in Lemma 2 and the fact that the weighting function is separable yields

$$
\Lambda_{\hat f} \;=\; \int w_f(x,y)\, \mathrm{d}\big[\hat F(x,y) - F(x,y)\big] - \mu_X \int w_f(y)\, \mathrm{d}\big[\hat F(y) - F(y)\big]
$$

$$
\begin{aligned}
&- \mu_Y \int w_f(x)\, \mathrm{d}\big[\hat{F}(x) - F(x)\big] + O\left(\big\|\hat{f}(x,y) - f(x,y)\big\|^2_{(2,m)}\right) \\
= \;& \frac{1}{T}\sum_{t=1}^{T} a' A_t + o_p\left(T^{-1/2}\right),
\end{aligned}
$$

where $a' = (1, -\mu_X, -\mu_Y)$. This means that $\sqrt{T}\,\Lambda_{\hat{f}} \xrightarrow{d} N(0, a'\Omega a)$ with

$a' E\big[A_t A'_{t+k}\big] a = \gamma_{XY}(k) + \gamma_X(k)\mu_X^2 + \gamma_Y(k)\mu_Y^2 - 2\big[\gamma_X(k) + \gamma_Y(k)\big]\mu_{XY}.$

Lastly, assumption A5 ensures that one may estimate consistently the above asymptotic variance using the tools provided by Newey and West (1987).

Proof of Proposition 2. The conditions imposed are such that the functional Taylor expansion holds even when both x_{tT} and y_{tT} are double arrays. Thus, it ensues that, under $H_{1,T}$ and assumptions A1 to A4,

$$
\hat{r}_q^w - \frac{T^{-1/2}\hat{\sigma}_w^{-1}}{1-q}\sum_{t=1}^{T} w_f(x_{tT}, y_{tT})\left\{1 - \left[\frac{f_X(x_{tT})f_Y(y_{tT})}{f_{XY}(x_{tT}, y_{tT})}\right]^{1-q}\right\} \xrightarrow{d^{[T]}} N(0,1),
$$

where the superscript $[T]$ denotes dependence on $f_{XY}^{[T]}$. The result then follows by noting that $\hat{\sigma}_w \xrightarrow{p^{[T]}} \sigma_w$ and

$$
\begin{aligned}
\Lambda_{f^{[T]}} &= \frac{1}{1-q} E\left\{ w_f(x_{tT}, y_{tT}) \left[1 - \left(\frac{f_X^{[T]}(x_{tT}) f_Y^{[T]}(y_{tT})}{f_{XY}^{[T]}(x_{tT}, y_{tT})}\right)^{1-q}\right]\right\} \\
&= E\Big[w_f(x_{tT}, y_{tT})\epsilon_T \lambda_{XY}(x_{tT}, y_{tT})\Big] = \epsilon_T\, \delta_\lambda.
\end{aligned}
$$

Proof of Corollary. It suffices to apply Theorem 1 and show that the asymptotic variance σ_w^2 reduces to the variance of the process implied by the weighting function. To appreciate this, notice that if $Y_t = X_{t-i}$, then $\mu_Y = \mu_X$, $\gamma_Y(k) = \gamma_X(k)$, and $\gamma_{XY}(k) = \tau_X^2(k) - \mu_X^4$. Further, serial independence implies that $\gamma_X(k) = 0$ for all $k \neq 0$, hence

$$
\begin{aligned}
\sigma_w^2 &= \gamma_{XY}(0) + 2\gamma_X(0)\mu_X^2 - 4\gamma_X(0)\mu_X^2 = \gamma_{XY}(0) - 2\gamma_X(0)\mu_X^2 \\
&= \tau_X^2(0) - \mu_X^4 - 2\left[\tau_X(0) - \mu_X^2\right]\mu_X^2 = \left[\tau_X(0) - \mu_X^2\right]^2 = \gamma_X^2(0).
\end{aligned}
$$

Proof of Theorem 2. Consider a model given by $Y_t = Y_t(\theta_0)$ and a T^d-consistent estimator $\hat{\theta}$ of θ_0. The interest lies on testing model specification by checking whether Y_t is serially independent, but Y_t is unobservable and the testing procedure must be carried out using $\hat{Y}_t = Y_t\big(\hat{\theta}\big)$. The test is nuisance parameter free if the statistic evaluated at $\hat{\theta}$, i.e.

$$\Lambda_{\hat{f}}(\hat{\theta}) = \frac{1}{(1-q)(T-i)} \sum_{t=i+1}^{T} w(\hat{\theta}) \left\{ 1 - \left[\frac{\hat{f}\left(Y_t(\hat{\theta})\right) \hat{f}\left(Y_{t-i}(\hat{\theta})\right)}{\hat{f}\left(Y_t(\hat{\theta}), Y_{t-i}(\hat{\theta})\right)} \right]^{1-q} \right\},$$

where $w(\theta) = w\big(Y_t(\theta), Y_{t-i}(\theta)\big)$, converges to the same distribution of the statistic evaluated at the true parameter vector θ_0, i.e. $\Lambda_{\hat{f}}(\theta_0)$. The limiting distribution derived in Theorem 1 applies to $\Lambda_{\hat{f}}(\theta_0)$, hence it is natural to pursue a second-order Taylor expansion with Lagrange remainder of $\Lambda_{\hat{f}}(\hat{\theta})$ about $\Lambda_{\hat{f}}(\theta_0)$, i.e.

$$\begin{aligned} \Lambda_{\hat{f}}(\hat{\theta}) &= \Lambda_{\hat{f}}(\theta_0) + \Lambda'_{\hat{f}}(\theta_0)\left(\hat{\theta} - \theta_0\right) + \frac{1}{2}\Lambda''_{\hat{f}}(\theta_*)\left(\hat{\theta} - \theta_0, \hat{\theta} - \theta_0\right) \\ &= \Lambda_{\hat{f}}(\theta_0) + \Delta_{1T} + \Delta_{2T}, \end{aligned}$$

where $\theta_* \in \big[\theta_0, \hat{\theta}\big]$ and $\Lambda'_{\hat{f}}$ and $\Lambda''_{\hat{f}}$ denote the first and second order differentials with respect to θ, respectively. Let $Z_t(\theta) = \big(Y_t(\theta), Y_{t-1}(\theta)\big)$ and $Z_t = \big(Y_t(\theta_0), Y_{t-1}(\theta_0)\big)$. The first differential reads

$$\begin{aligned} \Lambda'_f(\theta_0) &= \frac{1}{1-q} \int (\, w'_z f_z + w_z f'_z \,) \left[1 - (g_z/f_z)^{1-q}\right] \mathrm{d}z \\ &\quad + \int w_z \,(f_z/g_z)^q \,(\, g_z f'_z/f_z - g'_z \,)\, \mathrm{d}z \\ &= \frac{1}{1-q} \int (\, w'_z f_z + w_z f'_z \,) \left[1 - (g_z/f_z)^{1-q}\right] \mathrm{d}z \\ &\quad + \int w_z \,(f_z/g_z)^q \, g_z \, \log(f_z/g_z)' \, \mathrm{d}z, \end{aligned}$$

where all differentials are with respect to θ evaluated at θ_0. Since the kernel estimates of the density function and its derivative are such that $\left(\hat{f}_z - f_z\right)^2 = O_p\left(T^{-1} b_{z,T}^{-1}\right)$ and $\left(\hat{f}'_z - f'_z\right)^2 = O_p\left(T^{-1} b_{z,T}^{-3}\right)$, $\Lambda'_{\hat{f}}(\theta_0) = O_p\left(T^{-1} b_{z,T}^{-2}\right)$. Therefore, Δ_{1T} is of order $O_p\left(T^{-(1+d)} b_{z,T}^{-2}\right)$. The second term requires more caution for it is not evaluated at the true parameter θ_0. It is not difficult to show, however, that

$$\sup_{|\theta_* - \theta_0| < \epsilon} \left| \Lambda''_{\hat{f}}(\theta_*) \right| = O_p\left(T^{-1} b_{z,T}^{-3}\right),$$

which implies that Δ_{2T} is of order $O_p\left(T^{-(1+2d)}b_{z,T}^{-3}\right)$. The limiting distributions of $\Lambda_{\hat{f}}\left(\hat{\theta}\right)$ and $\Lambda_{\hat{f}}(\theta_0)$ then coincide if and only if $T^{1/2}(\Delta_{1T} + \Delta_{2T}) = o_p(1)$. As assumption A4 imposes that $b_{z,T} = b_{y,T}^2 = o\left(T^{-1/(s+1)}\right)$, it ensues that

$$T^{1/2}(\Delta_{1T} + \Delta_{2T}) = T^{1/2}\left[O_p\left(T^{-(1+d)}b_{z,T}^{-2}\right) + O_p\left(T^{-(1+2d)}b_{z,T}^{-3}\right)\right],$$

which is $o_p(1)$ for $d \geq \max\{2/(s+1) - 1/2, 3/(2s+2) - 1/4\}$.

Notes

1. Alternatively, one may assume normality to check independence through the cross-correlation function. For instance, Hong's (1996) develops a coherency-based test to check whether two covariance-stationary processes are uncorrelated by first prewhitening the time series and then gauging the sum of finitely many squares of residual cross-correlations.

2. Robinson (1991), Hong and White (2000) and Zheng (2000) assume that the densities are bounded to derive their asymptotic tests of serial independence based on the Kullback-Leibler entropy. The former further relies on a sample-splitting device to work out the asymptotic theory, whereas the latter two Taylor-expand the Kullback-Leibler measure to find the limiting distribution of their respective test statistics. Unfortunately, the solution by Taylor expansion given by Hong and White (2000) and Zheng (2000) does not seem applicable to entropic indexes different than $q \to 1$.

3. For the limit case ($q \to 1$) where the Tsallis entropy recovers the Kullback-Leibler information criterion, the local alternative of interest is simply $H_{1,T}: \ \sup_{(x,y)\in\mathcal{S}} \left| f_{XY}^{[T]}(x,y) - g_{XY}^{[T]}(x,y) \exp\left[\epsilon_T \lambda_{XY}(x,y)\right]\right|$.

4. Actually, randomizing the length of the blocks is not enough to guarantee the stationarity of the resampled time series. As the blocks overlap, the first and last original observations appear in fewer blocks than the rest. Therefore, to deal with these end effects, the stationary bootstrap wraps the data around a circle, so that X_1 follows X_T.

References

Aït-Sahalia, Y. (1994), The delta method for nonparametric kernel functionals, Graduate School of Business, University of Chicago.

Andrews, D. W. K. (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59**, 817–858.

Baek, E. G. and Brock, W. A. (1992), A nonparametric test for independence of a multivariate time series, *Statistica Sinica* **2**, 137–156.

Bosq, D. (1996), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, Springer-Verlag, New York.

Brock, W. A., Dechert, W. D., Scheinkman, J. A. and LeBaron, B. (1996), A test for independence based on the correlation dimension, *Econometric Reviews* **15**, 197–235.

Carlstein, E. (1986), The use of subseries values for estimating the variance of a general statistic from a stationary sequence, *Annals of Statistics* **14**, 1171–1179.

de Lima, P. J. F. (1996), Nuisance parameter free properties of correlation integral based statistics, *Econometric Reviews* **15**, 237–259.

Drost, F. C. and Werker, B. J. M. (1993), A note on Robinson's test of independence, *CentER Discussion Paper 9315*, Tilburg University.

Fan, Y. (1995), Bootstrapping a consistent nonparametric goodness-of-fit test, *Econometric Reviews* **14**, 367–382.

Fan, Y. and Li, Q. (1996), Consistent model specification tests: Omitted variables and semiparametric functional forms, *Econometrica* **64**, 865–890.

Fan, Y. and Linton, O. (1997), Some higher order theory for a consistent nonparametric model specification, University of Windsor and Yale University.

Gregory, A. W. and Sampson, M. J. (1991), Testing long-run properties of stationary time series, *Journal of Business and Economic Statistics* **9**, 287–295.

Hall, P. (1984), Central limit theorem for integrated squared error multivariate nonparametric density estimators, *Journal of Multivariate Analysis* **14**, 1–16.

Hall, P. (1985), Resampling a coverage process, *Stochastic Processes and Applications* **20**, 231–246.

Hallin, M. and Puri, M. L. (1992), Rank tests for time series analysis: A survey, *in* D. R. Brillinger, P. Caines, J. Geweke, E. Parzan, M. Rosenblatt and M. S. Taqqu (eds), *New Directions in Time Series Analysis*, Springer-Verlag, New York.

Hallin, M., Jurečková, J., Picek, J. and Zahaf, T. (1997), Nonparametric tests of independence of two autoregressive time series based on autoregression rank scores, *ISRO Série Probabilités et Statistique 81*, Université Libre de Bruxelles.

Hart, J. D. (1997), *Nonparametric smoothing and lack-of-fit tests*, Springer-Verlag, New York.

Hjellvik, V. and Tjøstheim, D. (1996a), Nonparametric statistics for testing of linearity and serial independence, *Journal of Nonparametric Statistics* **6**, 223–251.

Hjellvik, V. and Tjøstheim, D. (1996b), Nonparametric tests of linearity for time series, *Biometrika* **82**, 351–368.

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

Hong, Y. (1996), Testing for independence between two covariance-stationary time series, *Biometrika* **100000**, 33500–340000.

Hong, Y. and White, H. (2000), Asymptotic distribution theory for nonparametric entropy measures of serial dependence, Cornell University and Ynivlersity of California at San Diego.

Johnson, D. and McClelland, R. (1998), A general dependence test and applications, *Journal of Applied Econometrics* **13**, 627–644.

Khashimov, S. A. (1992), Limiting behaviour of generalized U-statistics of weakly dependent stationary processes, *Theory of Probability and Its Applications* **37**, 148–150.

Künsch, H. R. (1989), The jackknife and the bootstrap for general stationary observations, *Annals of Statistics* **17**, 1217–1241.

Liu, R. Y. and Singh, K. (1992), Moving blocks jacknife and bootstrap capture weak dependence, *in* R. LePage and L. Billard (eds), *Exploring the limits of bootstrap*, John Wiley, New York.

Mizrach, B. (1995), A simple nonparametric test for independence, Rutgers University.

Newey, W. K. and West, K. D. (1987), A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**, 703–708.

Phillips, P. C. B. (1991), Error correction and long-run equilibrium in continuous time, *Econometrica* **59**, 967–980.

Pinkse, J. (1994), *Nonparametric and semiparametric estimation and testing*, PhD thesis, London School of Economics and Political Science.

Pinkse, J. (1998), A consistent nonparametric test for serial independence, *Journal of Econometrics* **84**, 205–231.

Pinkse, J. (1999), Nonparametric misspecification testing, University of British Columbia.

Pitman, E. J. G. (1979), *Some Basic Theory for Statistical Inference*, Chapman Hall, London.

Politis, D. N. and Romano, J. P. (1994), The stationary bootstrap, *Journal of the American Statistical Association* **89**, 1303–1313.

Robinson, P. M. (1991), Consistent nonparametric entropy-based testing, *Review of Economic Studies* **58**, 437–453.

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

Skaug, H. J. and Tjøstheim, D. (1993), Nonparametric tests of serial independence, *in* T. S. Rao (ed.), *Developments in Time Series Analysis*, Chapman and Hall, London.

Skaug, H. J. and Tjøstheim, D. (1996), Measures of distance between densities with application to testing for serial independence, *in* P. Robinson and M. Rosenblatt (eds), *Time Series Analysis in Memory of E. J. Hannan*, Springer, New York.

Staniswalis, J. G. and Severini, T. A. (1991), Diagnostics for assessing regression models, *Journal of the American Statistical Association* **86**, 684–692.

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

Tsallis, C. (1988), Possible generalization of Boltzmann-Gibbs statistics, *Journal of Statistical Physics* **52**, 479–487.

Tsallis, C. (1998), Generalized entropy-based criterion for consistent testing, *Physical Review E* **58**, 1442–1445.

White, H. (1999), A reality check for data snooping, *Econometrica* **68**, 1097–1126.

Zheng, J. X. (2000), A consistent test of conditional parametric distributions, *Econometric Theory* **16**, 667–691.

Table 1

Finite sample properties of tests of independence

$H_0: \; X_t \perp\!\!\!\perp Y_t$, where

$$Y_t = 0.8\, Y_{t-1} + \upsilon_t, \qquad \upsilon_t \sim \text{ iid } N(0,1)$$

$$X_t = 0.8\, X_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0,1)$$

$T = 500$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.2765	0.3257	0.4315	0.7027
median		0.2669	0.3128	0.4110	0.6414
standard deviation		0.0753	0.0933	0.1376	0.2912
skewness		0.6980	0.8333	1.0731	2.0082
kurtosis		3.8873	4.3290	5.2054	12.1181
quantiles:	0.01	0.1364	0.1600	0.2052	0.2880
	0.05	0.1696	0.1953	0.2476	0.3636
	0.10	0.1871	0.2168	0.2765	0.4108
	0.25	0.2234	0.2599	0.3350	0.5054
	0.75	0.3210	0.3795	0.5046	0.8336
	0.90	0.3747	0.4467	0.6084	1.0626
	0.95	0.4106	0.4962	0.6844	1.2160
	0.99	0.5004	0.6030	0.8608	1.6428
$T = 1000$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.2481	0.2942	0.3910	0.6202
median		0.2438	0.2891	0.3764	0.5846
standard deviation		0.0598	0.0738	0.1080	0.2104
skewness		0.5357	0.6551	0.8623	1.2923
kurtosis		3.5074	3.7805	4.3087	5.9011
quantiles:	0.01	0.1326	0.1588	0.2075	0.3014
	0.05	0.1582	0.1866	0.2383	0.3576
	0.10	0.1730	0.2034	0.2635	0.3867
	0.25	0.2042	0.2412	0.3153	0.4707
	0.75	0.2839	0.3372	0.4507	0.7255
	0.90	0.3268	0.3896	0.5305	0.8924
	0.95	0.3540	0.4285	0.5854	1.0076
	0.99	0.4101	0.5025	0.7219	1.3353

Number of replications: 2000

Table 2

Finite sample properties of tests of independence

$H_0: \; X_t \perp\!\!\!\perp Y_t$, where

$$Y_t = 0.8\,Y_{t-1} + X_t + \upsilon_t, \quad \upsilon_t \sim \text{ iid } N(0,1)$$

$$X_t = 0.8\,X_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0,1)$$

$T = 500$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		2.8381	3.5284	5.4891	17.9120
median		2.8237	3.5065	5.4449	17.3345
standard deviation		0.3883	0.4759	0.8123	4.5945
skewness		0.1915	0.2054	0.2961	0.8941
kurtosis		3.0447	3.0766	3.2478	4.5746
quantiles:	0.01	1.9826	2.5021	3.7589	9.7432
	0.05	2.2259	2.7758	4.2120	11.3520
	0.10	2.3538	2.9303	4.4793	12.6164
	0.25	2.5765	3.2043	4.9389	14.7580
	0.75	3.0888	3.8309	6.0071	20.5612
	0.90	3.3433	4.1546	6.5306	23.8118
	0.95	3.4921	4.3356	6.8889	26.2527
	0.99	3.8255	4.7731	7.6214	32.2606
power:	0.01	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	1.0000	1.0000	1.0000
$T = 1000$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		3.9931	5.0824	8.1342	28.2330
median		3.9954	5.0806	8.1246	27.6460
standard deviation		0.4038	0.4977	0.8728	5.4770
skewness		0.1670	0.1443	0.1856	0.6636
kurtosis		2.9775	2.9512	3.0076	3.7629
quantiles:	0.01	3.1084	3.9642	6.2140	17.7252
	0.05	3.3283	4.2607	6.7126	20.4246
	0.10	3.4797	4.4521	7.0343	21.6816
	0.25	3.7183	4.7314	7.5121	24.4286
	0.75	4.2508	5.3935	8.6991	31.3367
	0.90	4.5157	5.7321	9.2671	35.4781
	0.95	4.6950	5.9428	9.6344	38.3175
	0.99	4.9663	6.2710	10.1615	43.7365
power:	0.01	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	1.0000	1.0000	1.0000

Number of replications: 2000

Table 3

Finite sample properties of tests of independence

$H_0: \; X_t \perp\!\!\!\perp Y_t$, where

$$Y_t = 0.8\,Y_{t-1} + X_t\,\upsilon_t, \quad \upsilon_t \sim \text{ iid } N(0,1)$$
$$X_t = 0.8\,X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{ iid } N(0,1)$$

$T = 500$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.7250	0.8556	1.1534	2.1738
median		0.7147	0.8400	1.1254	2.0164
standard deviation		0.1517	0.1844	0.2750	0.8263
skewness		0.5081	0.5365	0.6820	1.8125
kurtosis		3.6293	3.6525	3.9930	9.4436
quantiles:	0.01	0.4112	0.4868	0.6306	0.9765
	0.05	0.4973	0.5790	0.7516	1.1752
	0.10	0.5431	0.6323	0.8264	1.3351
	0.25	0.6230	0.7321	0.9663	1.6343
	0.75	0.8133	0.9650	1.3099	2.5071
	0.90	0.9219	1.0942	1.5125	3.1814
	0.95	0.9898	1.1840	1.6462	3.7918
	0.99	1.1596	1.3843	1.9716	5.0647
power:	0.01	0.9465	0.9290	0.8660	0.7445
	0.05	0.9905	0.9890	0.9775	0.9405
$T = 1000$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.8878	1.0726	1.4792	2.7467
median		0.8766	1.0587	1.4547	2.6223
standard deviation		0.1470	0.1832	0.2762	0.7463
skewness		0.4102	0.4185	0.4966	1.1205
kurtosis		3.6429	3.5428	3.5235	3.8653
quantiles:	0.01	3.4113	3.3555	3.4329	5.2334
	0.05	0.5781	0.6894	0.9222	1.5111
	0.10	0.6621	0.7916	1.0601	1.7475
	0.25	0.7081	0.8520	1.1530	1.9400
	0.75	0.7879	0.9490	1.2898	2.2354
	0.90	0.9833	1.1862	1.6423	3.1260
	0.95	1.0725	1.3075	1.8350	3.6923
	0.99	1.2708	1.5604	2.2388	5.0715
power:	0.01	1.0000	1.0000	1.0000	0.9950
	0.05	1.0000	1.0000	1.0000	1.0000

Number of replications: 2000

Table 4

Finite sample properties of tests of serial independence

$H_0: \; X_t \perp\!\!\!\perp X_{t-1}$, where $X_t \sim$ iid $N(0,1)$

$T = 500$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.8421	0.9696	1.2321	1.8056
median		0.8186	0.9445	1.1933	1.7248
standard deviation		0.3089	0.3414	0.4201	0.6318
skewness		0.4857	0.5363	0.6425	0.8602
kurtosis		3.5603	3.6516	3.8489	4.4166
quantiles:	0.01	0.2230	0.2959	0.4383	0.6786
	0.05	0.3828	0.4644	0.6209	0.9419
	0.10	0.4578	0.5496	0.7225	1.0675
	0.25	0.6246	0.7306	0.9299	1.3452
	0.75	1.0398	1.1818	1.4809	2.1846
	0.90	1.2439	1.4079	1.7814	2.6250
	0.95	1.3621	1.5506	1.9689	2.9410
	0.99	1.6562	1.8972	2.4158	3.7039
$T = 1000$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.8046	0.9293	1.1833	1.7227
median		0.7960	0.9151	1.1520	1.6581
standard deviation		0.2866	0.3129	0.3773	0.5470
skewness		0.3347	0.4172	0.5698	0.8027
kurtosis		3.3471	3.4971	3.7892	4.2802
quantiles:	0.01	0.2157	0.3085	0.4507	0.7256
	0.05	0.3672	0.4592	0.6252	0.9539
	0.10	0.4486	0.5416	0.7232	1.0879
	0.25	0.6082	0.7138	0.9184	1.3379
	0.75	0.9860	1.1252	1.4152	2.0400
	0.90	1.1734	1.3260	1.6691	2.4108
	0.95	1.3013	1.4711	1.8463	2.7358
	0.99	1.5581	1.7533	2.2326	3.3968

Number of replications: 2000

Table 5

Finite sample properties of tests of serial independence

$H_0: X_t \perp\!\!\!\perp X_{t-1}$, where $X_t \sim$ iid χ^2_1

$T = 500$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.8134	0.9289	1.1683	1.7016
median		0.7911	0.9034	1.1257	1.6119
standard deviation		0.2934	0.3217	0.3909	0.5825
skewness		0.3726	0.4295	0.5611	0.8443
kurtosis		3.1594	3.2716	3.5724	4.4582
quantiles:	0.01	0.2132	0.2823	0.4054	0.6662
	0.05	0.3694	0.4405	0.5891	0.8715
	0.10	0.4568	0.5456	0.7035	1.0297
	0.25	0.6075	0.7029	0.9023	1.3047
	0.75	0.9985	1.1264	1.4017	2.0342
	0.90	1.2106	1.3586	1.6982	2.4673
	0.95	1.3306	1.5066	1.8971	2.7622
	0.99	1.5933	1.7919	2.2291	3.3342
$T = 1000$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.7943	0.9106	1.1490	1.6640
median		0.7724	0.8831	1.1066	1.5997
standard deviation		0.2687	0.2929	0.3522	0.5113
skewness		0.3984	0.4544	0.5688	0.7708
kurtosis		3.2680	3.3877	3.6237	4.0467
quantiles:	0.01	0.2450	0.3195	0.4852	0.7174
	0.05	0.3823	0.4677	0.6227	0.9290
	0.10	0.4617	0.5493	0.7303	1.0777
	0.25	0.6071	0.7095	0.9046	1.3114
	0.75	0.9671	1.0932	1.3626	1.9577
	0.90	1.1482	1.2867	1.6057	2.3352
	0.95	1.2586	1.4191	1.7686	2.5916
	0.99	1.4989	1.6957	2.1435	3.1778

Number of replications: 2000

Table 6

Finite sample properties of tests of serial independence

$H_0: \ X_t \perp\!\!\!\perp X_{t-1}$, where $X_t \sim$ iid $t(5)$

$T = 500$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.3796	0.4463	0.5894	0.9434
median		0.3606	0.4199	0.5437	0.8341
standard deviation		0.2490	0.2708	0.3320	0.5825
skewness		0.3935	0.4965	0.8289	2.4623
kurtosis		3.1368	3.3648	4.5172	17.8770
quantiles:	0.01	-0.1262	-0.0867	-0.0059	0.1186
	0.05	0.0058	0.0494	0.1243	0.2648
	0.10	0.0765	0.1209	0.2082	0.3672
	0.25	0.2011	0.2520	0.3520	0.5625
	0.75	0.5435	0.6202	0.7871	1.1817
	0.90	0.6995	0.7972	1.0267	1.6259
	0.95	0.8256	0.9303	1.1674	1.9770
	0.99	1.0542	1.1992	1.5680	2.8352
$T = 1000$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		0.3887	0.4554	0.5953	0.9203
median		0.3817	0.4486	0.5747	0.8571
standard deviation		0.2346	0.2503	0.2945	0.4577
skewness		0.3465	0.4044	0.5934	1.3428
kurtosis		3.4678	3.5769	3.9700	6.7642
quantiles:	0.01	-0.1025	-0.0461	0.0268	0.1481
	0.05	0.0288	0.0698	0.1568	0.3243
	0.10	0.1031	0.1504	0.2463	0.4305
	0.25	0.2180	0.2745	0.3830	0.5973
	0.75	0.5450	0.6219	0.7742	1.1530
	0.90	0.6829	0.7728	0.9709	1.4675
	0.95	0.7832	0.8818	1.1037	1.7141
	0.99	0.9843	1.1199	1.4355	2.4583

Number of replications: 2000

Table 7

Finite sample properties of tests of serial independence

$H_0: \; X_t \perp\!\!\!\perp X_{t-1}$, where

$$X_t = 0.8\, X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{ iid } N(0,1)$$

$T = 500$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		14.2611	17.2157	24.7041	55.9007
median		14.1964	17.0838	24.4742	54.9792
standard deviation		2.1253	2.5272	3.8565	11.5334
skewness		0.2046	0.2221	0.2635	0.4105
kurtosis		3.0806	3.1183	3.1547	3.2676
quantiles:	0.01	9.7655	11.8736	16.7538	33.0857
	0.05	10.8803	13.2214	18.7740	38.5277
	0.10	11.6189	14.1145	19.9015	41.5589
	0.25	12.7653	15.4414	22.0183	47.6305
	0.75	15.6961	18.9402	27.2574	63.2438
	0.90	17.0534	20.5151	29.6456	70.9626
	0.95	17.6734	21.3669	31.2170	75.7506
	0.99	19.6283	23.5696	34.3304	86.4352
power:	0.01	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	1.0000	1.0000	1.0000
$T = 1000$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		20.2433	25.0056	36.8842	87.5425
median		20.1335	24.8565	36.6769	86.6119
standard deviation		2.2474	2.6648	4.1397	13.2489
skewness		0.1820	0.1657	0.1851	0.3440
kurtosis		3.1585	3.1966	3.2280	3.3074
quantiles:	0.01	15.1429	18.7529	27.3509	58.2691
	0.05	16.6620	20.7925	30.3899	67.2600
	0.10	17.4835	21.7495	31.8896	71.5537
	0.25	18.7291	23.2329	34.1124	78.3341
	0.75	21.6578	26.7252	39.5757	95.7403
	0.90	23.1294	28.4485	42.1843	105.0111
	0.95	24.1828	29.5713	43.9969	110.3170
	0.99	25.9606	31.6917	47.0980	122.4792
power:	0.01	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	1.0000	1.0000	1.0000

Number of replications: 2000

Table 8

Finite sample properties of tests of serial independence

$H_0: \ X_t \perp\!\!\!\perp X_{t-1}$, where

$$X_t = \left(0.2 + 0.8\, X_{t-1}^2\right)^{1/2} \epsilon_t, \quad \epsilon_t \sim \text{ iid } N(0,1)$$

$T = 500$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		4.5610	5.0445	6.0841	10.1137
median		4.3250	4.7841	5.7470	8.0739
standard deviation		1.4612	1.5736	1.9430	31.0409
skewness		0.9331	0.9394	1.1886	31.9183
kurtosis		4.1423	4.1949	5.3785	1102.2963
quantiles:	0.01	2.0836	2.4089	2.9617	3.9785
	0.05	2.6373	2.9601	3.6068	4.8539
	0.10	2.9344	3.2770	3.9412	5.4057
	0.25	3.4912	3.9040	4.7053	6.4620
	0.75	5.3927	5.9316	7.0654	10.1638
	0.90	6.4889	7.0909	8.5217	13.0296
	0.95	7.3188	8.0296	9.6822	15.8050
	0.99	8.9919	9.8889	12.4928	29.6317
power:	0.01	0.9995	1.0000	0.9995	0.9995
	0.05	1.0000	1.0000	1.0000	1.0000
$T = 1000$		$q = 1/2$	$q \to 1$	$q = 2$	$q = 4$
mean		6.3411	7.1534	8.9172	43.5420
median		5.9606	6.7358	8.3140	12.0287
standard deviation		2.0393	2.2444	3.4130	1213.1804
skewness		1.8281	1.9451	6.7246	44.4974
kurtosis		9.6112	10.9394	108.3416	1986.4580
quantiles:	0.01	3.3221	3.8195	4.8308	6.8104
	0.05	3.8941	4.5079	5.5838	7.7427
	0.10	4.3075	4.9136	6.0967	8.5682
	0.25	4.9787	5.6572	6.9925	9.9362
	0.75	7.2011	8.0838	9.9846	14.9448
	0.90	8.8265	9.8599	12.1904	19.2823
	0.95	10.0072	11.1850	13.7697	23.1523
	0.99	13.3811	14.8042	18.9399	62.0778
power:	0.01	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	1.0000	1.0000	1.0000

Number of replications: 2000

Table 9

Empirical size of bootstrap-based tests at the 5% level of significance

$H_0: \; X_t \perp\!\!\!\perp Y_t$, where

$$Y_t = \phi\, Y_{t-1} + \upsilon_t, \qquad \upsilon_t \sim \text{ iid } N(0,1)$$

$$X_t = \phi\, X_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0,1)$$

sample size	index	autoregressive coefficient ϕ						
		0.4	0.5	0.6	0.7	0.8	0.9	0.95
$T = 500$	$q = 1/2$	0.074	0.086	0.084	0.088	0.092	0.108	0.136
	$q \to 1$	0.080	0.084	0.090	0.094	0.106	0.118	0.164
	$q = 2$	0.110	0.122	0.142	0.164	0.168	0.184	0.256
	$q = 4$	0.146	0.168	0.188	0.232	0.256	0.288	0.368
$T = 1000$	$q = 1/2$	0.064	0.076	0.084	0.088	0.088	0.096	0.126
	$q \to 1$	0.072	0.074	0.084	0.088	0.100	0.108	0.142
	$q = 2$	0.116	0.122	0.140	0.162	0.178	0.186	0.234
	$q = 4$	0.128	0.150	0.168	0.194	0.226	0.262	0.326

Number of replications: 500 Number of bootstrap samples: 99