Nonparametric entropy-based tests of independence between stochastic processes

Marcelo Fernandes
Queen Mary, University of London
m.fernandes@qmul.ac.uk

Breno Néri
New York University
breno.neri@nyu.edu

Abstract: This paper develops nonparametric tests of independence between two stochastic processes satisfying β-mixing conditions. The testing strategy boils down to gauging the closeness between the joint and the product of the marginal stationary densities. For that purpose, we take advantage of a generalized entropic measure so as to build a whole family of nonparametric tests of independence. We derive asymptotic normality and local power using the functional delta method for kernels. As a corollary, we also develop a class of entropy-based tests for serial independence. The latter are nuisance parameter free, and hence also qualify for dynamic misspecification analyses. We then investigate the finite-sample properties of our serial independence tests through Monte Carlo simulations. They perform quite well, entailing more power against some nonlinear AR alternatives than two popular nonparametric serial-independence tests.

JEL classification numbers: C12, C14.
Keywords: independence, misspecification testing, nonparametric theory, Tsallis entropy.

Acknowledgments: We are indebted to Yacine Aït-Sahalia, Christophe Croux, Renato Flôres, Han Hong, Esfandiar Maasoumi (Editor), Marcelo Medeiros, Benedikt Pötscher, four anonymous referees, and seminar participants at the Université Libre de Bruxelles, the European University Institute, the 4th Spring Meeting of Young Economists (Amsterdam, 1999), and the Colóquio Brasileiro de Matemática (Rio de Janeiro, 2001) for valuable comments. Fernandes acknowledges with gratitude the financial support from the ESRC under the grant RES-062-23-0311. The usual disclaimer applies.

1 Introduction

Independence is one of the most valuable concepts in econometrics as most specification 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), Tjøstheim (1996), Pinkse (1998), Maasoumi and Racine (2002), Granger, Maasoumi and Racine (2004), and Racine and Maasoumi (2007).

The fact that stochastic processes are potentially serially dependent complicates the task of developing a suitable test. Consider a bivariate stochastic processes $\{(X_t, Y_t),\ t \geq 0\}$. The null hypothesis of interest then is

$$\mathbb{H}_0^* : \ f_{XY}(x_1, x_2, \ldots, y_1, y_2, \ldots) = f_X(x_1, x_2, \ldots) f_Y(y_1, y_2, \ldots) \ \text{ a.s.}$$

It is of course infeasible to develop a test without imposing additional structure. For instance, Haugh (1976), Geweke (1981) and Hong (1996) assume that $\{(X_t, Y_t),\ t \geq 0\}$ is a stationary Gaussian autoregressive process. Testing would then consist of two steps: One first obtains white noise residuals by performing univariate autoregressions, and then check the sample cross-correlations of the residuals.

To avoid imposing restrictive parametric assumptions such as linearity and normality, we take a different avenue. For the sake of argument, suppose that the joint process $\{(X_t, Y_t),\ t \geq 0\}$ does not feature any serial (cross-)dependence. One could then test the independence between $\{X_t\}$ and $\{Y_t\}$ by checking whether

$$\mathbb{H}_0 : \ f_{XY}(x_t, y_t) = f_X(x_t) f_Y(y_t) \ \text{ a.s.} \tag{1}$$

holds. Yet, even in the more general setting where $\{(X_t, Y_t),\ t \geq 0\}$ is a stationary stochastic process, 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 $\{(X_t, Y_t),\ t \geq 0\}$, and hence (1) corresponds to the property of long-run (or unconditional) independence. See Gregory and Sampson (1991) for further clarifications.

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 marginal densities of X_t and X_{t-i} as measured by the Kullback-Leibler information. Skaug and Tjøstheim (1993, 1996) extend Robinson's framework to other measures of discrepancy between densities, including the Hellinger distance. Maasoumi and Racine (2002) and Granger et al. (2004) consider tests of serial independence based on a normalization of the Hellinger distance, whereas

Racine and Maasoumi (2007) employ the latter for testing also other hypotheses, such as time reversibility and symmetry.

Alternative procedures include tests that examine restrictions on the joint characteristic function (Pinkse, 1998, 2000) and on the joint correlation integral (Baek and Brock, 1992; Brock, Dechert, Scheinkman and LeBaron, 1996). Actually, it is straightforward to show that correlation integrals approximate well certain mutual information measures (Diks and Manzan, 2002). These tests are particularly interesting for diagnostic checking purposes because they are nuisance parameter free and hence also applicable to residuals (see also de Lima, 1996; Racine and Maasoumi, 2007). Another valuable alternative approach relies on rank-based tests (see Hallin and Puri, 1992; Diks and Panchenko, 2007).

This paper proposes tests for the long-run (or unconditional) independence between two stationary stochastic processes satisfying β-mixing conditions. The null hypothesis of interest is as $\mathbb{H}_0$ in (1). The strategy relies on measuring the closeness between kernel estimates of the joint (unconditional) density and the product of the marginal (unconditional) densities of the processes. Instead of the conventional Euclidean distance, we 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 the usual quadratic distances (Skaug and Tjøstheim, 1996).

We thus extend the asymptotic results of Robinson (1991), Skaug and Tjøstheim (1996), Maasoumi and Racine (2002) and Granger et al. (2004) into two directions. First, we derive tests for the independence between stochastic processes rather than assuming serial independence under the null. We do that by focusing on the unconditional density of stochastic processes that satisfy a β-mixing condition. Second, we deal with a whole family of entropic measures rather than addressing only the particular cases of the Kullback-Leibler and Hellinger entropic measures. Although there are sound theoretical arguments in favor of the Hellinger distance (Maasoumi and Racine, 2002; Granger et al., 2004; Racine and Maasoumi, 2007), we show through Monte Carlo simulations that it pays off to hold a more agnostic view. The entropic index that entails the most powerful test indeed seem to vary according to the form of dependence.

Finally, we apply our testing procedures to check the residuals from a predictive regression between implied and realized volatility. In particular, we regress the monthly S&P 500 realized volatility on the previous options-implied volatility as measured by the VIX index. To assess unbiasedness, one should test whether the intercept and slope are respectively equal to zero and one, accounting for measurement error (Christensen and Prabhala, 1998), time-varying volatility premium (Chernov, 2007), and fractional integration (Bandi and Perron, 2006). Instead of focusing on unbiasedness, we take one step back and test only whether serial independence holds for the residuals. Our motivation rests on the fact that serial independence essentially rules out the presence of a persistent time-varying volatility premium. Our findings evince some serial dependence in the residuals, suggesting that the volatility premium may indeed vary over time.

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 we have in mind and provides asymptotic justification. In addition, we also demonstrate that our testing procedures are nuisance parameter free, and so suitable to specification testing. Section 4 then investigates the finite-sample properties of our independence tests through Monte Carlo simulations. The results are altogether very promising in that our entropy-based tests seem to compete well with the BDS test and to outperform the Hong and White's (2005) serial independence test based on a Taylor expansion of the Kullback-Leibler contrast. Section 5 then provides an empirical application in which we evaluate whether the residuals from from the predictive regression between implied and realized volatility are serial independent. Section 6 summarizes the main results and offers some concluding remarks, whereas the Appendix collects all technical lemmas and proofs.

2 Generalized entropic measure

Tsallis (1988) proposes the generalized entropy

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

where q denotes an entropic index different from one. In the limiting case $q \to 1$, the Tsallis entropy recovers the Kullback-Leibler information

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

whereas $q = 1/2$ implies that

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

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 (Hart, 1997; Golan, 2002).

Tsallis (1998) shows that, if the entropic index is positive, $\rho_q(f,g)$ is nonnegative, with equality holding if and only if f coincides with g almost everywhere. In addition, the Tsallis entropy satisfies another desirable property within the context of independence testing, namely, invariance under variable transformation in that $\rho_q[f(u),g(u)] = \rho_q[f_\ell(v),g_\ell(v)]$ for $v=\ell(u)$. Unfortunately, kernel-based estimates of the Tsallis entropy do not completely inherit this invariance property in view that kernel implementation is invariant only to affine transformations. In a similar vein, although Tsallis (1998) argues that it suffices to consider a family of entropy-based tests with $q \geq 1/2$ in view that $\frac{1}{q}\,\rho_q(f,g) = \frac{1}{1-q}\,\rho_{1-q}(g,f)$, this is not necessarily the case in practice for one cannot in general swap the kernel estimators of the two densities.[1] We thus derive the asymptotic theory for entropy-based testing procedures based on any positive entropic index.

There is an intimate connection between the entropy in (2) and other general families of information measures. For instance, Tsallis (1988) derive the Renyi's (1970) entropy as monotonically increasing function of (2), whereas Golan (2002) show a similar result relating the Chernoff, Cressie-Read, Renyi, and Tsallis measures (see also Chernoff, 1952; Cressie and Read, 1984; Golan and Perloff, 2002). In turn, Maasoumi (2002) and Granger et al. (2004) consider the very similar asymmetric k-class entropy given by

$$S_k(f,g) \equiv \frac{1}{k-1}\left\{\int [g(u)/f(u)]^k f(u)\mathrm{d}u - 1\right\}, \tag{5}$$

where the order k plays a similar role to the entropic index in (2). The above family indeed differs from the Tsallis and Renyi entropic measures only through their branching properties given that $\rho_q(f,g)$ and $S_k(g,f)$ are equivalent up to a monotonic decreasing transformation of the entropic index. See Golan (2002) and Golan and Perloff (2002) for further discussion about the relationship between these entropic measures.

Although Maasoumi and Racine (2002) and Granger et al. (2004) initially consider the above asymmetric k-class entropy, they focus their statistical analyses on a particular normalization of the Hellinger distance by setting $k=1/2$. Their motivation lies on the fact that the latter is a proper measure of distance given that it is symmetric and satisfies the triangular inequality. Although this is indeed a valid ex-ante criterion to assess the relative advantages of the different entropic measures, we take a more agnostic approach by investigating whether there is an entropic index

[1] We thank an anonymous referee for calling our attention to that.

that uniformly optimizes size and power within the family of kernel-based tests resulting from the Tsallis entropy. Accordingly, the next section investigates the asymptotic behavior of statistics based on (2) for any positive entropic index within the context of independence testing.

3 Asymptotic tests of independence

In what follows, we assume that $\{X_t,\ t \geq 0\}$ and $\{Y_t,\ t \geq 0\}$ are univariate processes with discretely recorded observations $(X_1, \ldots, X_T)$ and $(Y_1, \ldots, Y_T)$, respectively. We restrict attention to bivariate processes because of the curse of dimensionality that usually haunts any nonparametric density estimation. It is nonetheless straightforward to extend our results to higher-dimensional processes, especially if one employs higher-order kernels to alleviate the bandwidth conditions. We impose the following regularity conditions.

A1 $\{(X_t, Y_t),\ t \in \mathbb{N}\}$ is strictly stationary and $\beta-$mixing with $\beta_\tau = O\left(\rho^\tau\right)$, where $0 < \rho < 1$.

A2 The density functions f_{XY}, f_X, and f_Y are continuously differentiable up to the sth order and their derivatives are bounded and square integrable. In addition, the joint density function of $(Z_{k_1}, \ldots, Z_{k_\varsigma})$ with $Z_t = (X_t, Y_t)$, exists and satisfies the following Lipschitz-type condition:

$$|f(z_{k_1} + \Delta, \ldots, z_{k_\varsigma} + \Delta) - f(z_{k_1}, \ldots, z_{k_\varsigma})| \leq D(z_{k_1}, \ldots, z_{k_\varsigma})\, \|\Delta\|,$$

where D is integrable and $1 \leq \varsigma \leq 4$.

A3 Let $e_K \equiv \int K^2(u)\,\mathrm{d}u$ and $v_K \equiv \int \left[\int K(u)K(u+v)\,\mathrm{d}u\right]^2 \mathrm{d}v$, where the univariate kernel function K is symmetric around zero and of order s (even integer). We also assume that K is continuously differentiable up to the sth order on $\mathbb{R}$ with derivatives in $L^2\left(\mathbb{R}\right)$.

A4 (i) The bandwidths $b_{x,T}$ and $b_{y,T}$ are such that $b_{\cdot,T} \to 0$ and $T\, b_{\cdot,T} \to \infty$ as $T \to \infty$.
(ii) In addition, we also assume that both bandwidths are of order $o\left(T^{-1/(2s+1)}\right)$.

Assumption A1 restricts the amount of data dependence, requiring that the stochastic process is absolutely regular with geometric decay rate. Alternatively, one could assume $\alpha-$mixing conditions as in Gao and King (2004). See Chen, Linton and Robinson (2001) for some advantages of the $\beta-$mixing assumption relative to the $\alpha-$mixing condition in the context of kernel density estimation. Assumption A2 requires that the joint density f_{iXj} is smooth enough to admit a functional Taylor expansion. Assumption A3 focuses on higher-order kernels so as to reduce the bias in the kernel density estimation, thus alleviating, e.g., the bandwidth rates in Assumption A4(ii). The latter rates entail a slight degree of undersmoothing that eliminates some otherwise non-negligible bias terms in Proposition 1 (see Härdle and Mammen, 1993).

To test the null hypothesis $\mathbb{H}_0$, we 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

$$\hat{\rho}_q = \frac{1}{1-q} \left\{ 1 - \frac{1}{T} \sum_{t=1}^{T} \left[\frac{\hat{f}_X(X_t)\hat{f}_Y(Y_t)}{\hat{f}_{XY}(X_t, Y_t)} \right]^{1-q} \right\}. \tag{6}$$

The corresponding functional is

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

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 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. *Under Assumptions A1 to A4,*

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

where $\delta_{xy} = e_K^2 \int_{\mathcal{S}_{XY}} d(x,y)$, $\delta_x = e_K \int_{\mathcal{S}_X} dx$, $\delta_y = e_K \int_{\mathcal{S}_Y} dy$, *and* $\sigma^2 = v_K^2 \int_{\mathcal{S}_{XY}} d(x,y)$.

As is apparent, the asymptotic mean and variance exist only if the supports are bounded.[2] To avoid such a restrictive assumption, it is necessary to contrive some sort of weighting scheme. Consider next the following functional

$$\Lambda_f^w = \frac{1}{1-q} \int w_f(x,y) \left\{ 1 - \left[\frac{g_{XY}(x,y)}{f_{XY}(x,y)} \right]^{1-q} \right\} f_{XY}(x,y)\, \mathrm{d}(x,y), \tag{8}$$

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 (8), i.e.

$$\hat{\rho}_q^w = \frac{1}{T(1-q)} \sum_{t=1}^{T} w_f(X_t, Y_t) \left\{ 1 - \left[\frac{\hat{f}_X(X_t)\hat{f}_Y(Y_t)}{\hat{f}_{XY}(X_t, Y_t)} \right]^{1-q} \right\}, \tag{9}$$

one additional assumption is necessary.

[2] Robinson (1991), Hong and White (2005) and Zheng (2000) assume that the density supports are bounded to derive their asymptotic tests of serial independence based on the Kullback-Leibler entropy. Robinson further relies on a sample-splitting device to work out the asymptotic theory, whereas Hong and White and Zheng derive a Taylor-series expansion for the Kullback-Leibler measure. Unfortunately, the solution by Taylor expansion does not seem applicable to other entropic indexes.

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

$$(a)\ \mathbb{E}\left[w_f^{3+r}(x,y)\right] < \infty, \text{ for some } \epsilon \text{ such that } r > (1+3/\epsilon)(3+\epsilon/2),$$

$$(b)\ \mathbb{E}\left[\sup_{f^*\in N_f} w_{f^*}^2(x,y)\right] < \infty,$$

$$(c)\ \mathbb{E}\left[w_{f^*}(x,y) - w_{f^+}(x,y)\right]^2 \leq c\,\|f^* - f^+\|^2,$$

where c is a constant and $\|\cdot\|$ is the sup norm on $\mathbb{R}^2$.

Assumption A5 guarantees that one may truncate the infinite sum that appears in the asymptotic variance of the test statistic. In particular, a trimming function $w(x,y) = 1\!\!1_{\mathcal{S}}(x,y)$, where $\mathcal{S} = \mathcal{S}^X \times \mathcal{S}^Y$ is a compact subset of the density support, satisfies A5.

Before stating the next result, it is useful to establish some notation. Let $\mu_u = \mathbb{E}[w_f(u_t)]$, $\tau_u(k) = \mathbb{E}[w_f(u_t)w_f(u_{t+k})]$, 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 A4(i) and A5,*

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

where $\hat{\sigma}_w^2$ is any consistent estimator of the long-run variance

$$\sigma_w^2 = \sum_{k=-\infty}^{\infty} \left\{\gamma_{XY}(k) + \gamma_X(k)\mu_X^2 + \gamma_Y(k)\mu_Y^2 - 2[\gamma_X(k) + \gamma_Y(k)]\mu_{XY}\right\}.$$

To estimate consistently σ_w^2, Aït-Sahalia (1994, Theorem 4) shows that one may use the counterpart for the nonparametric kernel estimator of the one proposed by Newey and West (1987) for parametric estimators. The choice of the truncation lag is subject to the same provisions and improvements of the original Newey-West estimator (see Andrews, 1991). For instance, using the simple Bartlett kernel in the long-run variance estimator requires a truncation lag set at $O(n^{1/3})$, whereas it must be of order $O(n^{1/5})$ for the Parzen kernel.

Ergo, a test which rejects the null hypothesis at the level α when $\hat{r}_q^w$ 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

$$\mathbb{H}_{1,T} : \sup_{(x,y)\in\mathcal{S}} \left| f_{XY}^{[T]}(x,y) - g_{XY}^{[T]}(x,y)[1 + (q-1)\epsilon_T\lambda_{XY}(x,y)]^{\frac{1}{q-1}} \right| = o(\epsilon_T), \tag{10}$$

where $\epsilon_T = T^{-1/2}$ and λ_{XY} is such that $\delta_\lambda = \mathbb{E}[w_f(x,y)\lambda_{XY}(x,y)]$ exists. For the limit case ($q \to 1$) in which Tsallis entropy recovers the Kullback-Leibler information criterion, the term within bracket in (10) becomes $\exp\left[\epsilon_T \lambda_{XY}(x,y)\right]$.

Proposition 2. *Under Assumptions A1 to A4(i) and A5, the asymptotic local power is given by*

$$\Pr\left(\hat{r}_q^w \geq z_{1-\alpha} | \mathbb{H}_{1,T}\right) \longrightarrow 1 - \Phi\left(z_{1-\alpha} - \delta_\lambda / \sigma_w\right).$$

Unfortunately, the asymptotic local powers obtained by tests based on different entropic indexes cannot be directly compared since the local alternatives become closer to the null as q increases. This suggests that it may pay off to establish asymptotic results for a test that integrates across a range of entropic indexes.[3] This is however beyond the scope of this paper and so we defer it for further research.

How to select the weighting scheme is a difficult 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

$$w_t(x,y) = \begin{cases} \mathbb{1}_{\mathcal{S}}(x,y)(1+\gamma) & \text{if } t \text{ is odd} \\ \mathbb{1}_{\mathcal{S}}(x,y)(1-\gamma) & \text{if } t \text{ is even.} \end{cases} \tag{11}$$

As the latter design seems to produce tests with low power against both fixed (Drost and Werker, 1993) and local alternatives (Hong and White, 2005), we 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.

$$\Pr(X_0, \ldots, X_t) = \Pr(X_0) \ldots \Pr(X_t) \quad \text{a.s.} \tag{12}$$

[3] This is similar in spirit to Kočenda's (2001) method to eliminate the dependence of the BDS test on the proximity parameter by integrating across the correlation integral.

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

$$\mathbb{H}_0^i : f(X_t, X_{t-i}) = f(X_t) f(X_{t-i}) \text{ a.s.}, \tag{13}$$

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 $\mathbb{H}_0^i$ at the level α when $\sqrt{T}\,\hat{\rho}_{q,i}^w \geq z_{1-\alpha}\,\hat{\gamma}_X(0)$, where $\hat{\gamma}_X^2(0)$ is a consistent estimator of $\gamma_X^2(0) = \mathrm{var}[w_f(X_t)]$ and

$$\hat{\rho}_{q,i}^w = \frac{1}{(1-q)(T-i)} \sum_{t=i+1}^{T} w_f(X_t, X_{t-i}) \left\{ 1 - \left[\frac{\hat{f}(X_t)\hat{f}(X_{t-i})}{\hat{f}(X_t, X_{t-i})} \right]^{1-q} \right\} \tag{14}$$

is locally strictly unbiased.

Corollary. *Under Assumptions A1 to A4(i) and A5,*

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

where $\hat{\gamma}_X^2(0)$ is a consistent estimator of $\gamma_X^2(0) = var[w_f(X_t)]$.

The simplicity of the pairwise approach comes at the expense of an uncomfortable dependence on lags. Yet, one can mitigate this dependence by allowing for a null hypothesis such as $\mathbb{H}_0^s : \cap_{n=1}^N \mathbb{H}_0^{i_n}$ $(i_1 < \ldots < i_N)$ as in Skaug and Tjøstheim (1996). It is indeed possible to demonstrate that the partial sum statistic $\hat{\rho}_{q,w}^s = \sqrt{T} \sum_{n=1}^N \hat{\rho}_{q,i_n}^w$ 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. 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 $\hat{Y}_t = Y_t(\hat{\theta})$, $t = 1, \ldots, T$.

The next result establishes the conditions under 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 A4(i) and A5,*

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

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 becomes less stringent as the order of the kernel increases, because 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 Theorem 2 requires little.

4 Finite sample properties

It is well known that the asymptotic behavior of kernel-based tests is of little value in finite samples (see Fan and Linton, 2003) and that their results may heavily depend on the bandwidth (Skaug and Tjøstheim, 1993). We thus consider either permutation or bootstrap-based variants of our entropy-based tests by resampling under the null hypothesis of independence.[4] In particular, we run Monte Carlo simulations to assess the finite-sample performance of our serial independence tests relative to two alternative nonparametric tests. We inspect Hong and White's (2005) serial independence test based on a Taylor expansion of the Kullback-Leibler contrast (henceforth HW test) for it is arguably our main entropy-based competitor. In addition, we also employ the popular BDS test for it enjoys power against a wide array of alternatives (Brock et al., 1996).[5] To ensure a fair comparison, we focus on either permutation or bootstrap-based tests so as to alleviate any eventual finite-sample size distortion. In particular, we employ permutation for serial independence tests applied to raw data, whereas we bootstrap serial independence tests applied to residuals. All results rest on 1,000 Monte Carlo replications and bootstrap/permutation artificial samples.

We check exclusively for serial dependence of first order, at the 5% level of significance, allowing

[4] In the more general context of testing for serial independence between two stochastic processes, one must consider resampling procedures that are able to cope with serial dependence. Alternative solutions include, for instance, subsampling and block-bootstrap techniques, which are consistent even without any recentering of the test statistic.

[5] To maximize power, we compute the BDS test statistic with embedding dimension $m = 2$ and proximity parameter set to κ standard deviations, with $\kappa \in \{1/2, 1, 2, 4\}$. Because the quantitative results are very robust to variations in κ, we discuss only the size and power of the BDS test with $\kappa = 2$.

for two sample sizes. To avoid initialization problems, we construct samples of 500 and 1,000 observations for each data generating process by burning the first 500 realizations. For simplicity, we compute the entropy-based tests using simple trimming functions $w(x, y) = \mathbb{1}_{\mathcal{S}}(x, y)$ that allocate weight zero to observations out of the compact set $\mathcal{S} = \mathcal{S}^X \times \mathcal{S}^Y$. We consider two types of trimming functions. The first considers a symmetric support $\mathcal{S}_1^u = \{u : |u - \bar{u}| \leq \kappa_1 \hat{s}_u\}$ with $\kappa_1 \in \{1, 3/2, 2\}$, where $\bar{u}$ and $\hat{s}_u$ denote the sample mean and standard deviation, respectively. The second set trims observations out of a compact support $\mathcal{S}_2^u = \{u : q_u(\kappa_2) \leq u \leq q_u(1 - \kappa_2)\}$, where $q_u(\kappa_2)$ denotes the κ_2-quantile of the empirical distribution for $\kappa_2 \in \{0.10, 0.15, 0.20\}$. We carry out all density estimations using a Gaussian kernel and Silverman's (1986) rule-of-thumb bandwidth choice.

To examine size, we rely on a standard Gaussian noise specification, where $x_t = \epsilon_t \sim$ iid $N(0, 1)$. In addition, we also apply the testing procedures to the residuals of AR(1) processes so as to illustrate a context of misspecification testing. This is the only instance in which we employ bootstrap rather than permutation-based tests. We address power by considering different forms of serial dependence. Table 1 summarizes the data generating mechanisms that we contemplate. These include linear and nonlinear variants of the first-order AR process, a nonlinear moving average (MA), and ARCH and threshold GARCH processes. The nonlinear MA is a challenging alternative in that it involves serial dependence up to the second lag despite the fact we only employ first-order test statistics. We draw realizations of each data generating process for two sets of parameter values: either $(\alpha_0, \alpha_1, \alpha_2)$ is equal to $(0.1, 0.2, 0.5)$ or to $(0.1, 0.5, 0.5)$. The motivation for varying the value of α_1 lies on the fact that it dictates the strength of serial dependence.

Tables 2 and 3 report the results concerning the first trimming function with $\kappa_1 = 2$ for the sample sizes of 500 and 1,000 observations, respectively. As to what concerns size, the finite-sample performance of our entropy-based tests seems pretty competitive. This is of course not surprising in view that we employ resampling techniques (either permutation or bootstrap) to deal with any finite-sample size distortion. Despite the comparable magnitude in size distortions, our testing procedures are conservative displaying correct level, whereas the empirical sizes of the BDS and HW tests seem somewhat in excess. This result is a bit more salient within the context of misspecification testing, which involves applying the serial independence tests to AR(1) residuals. As expected, size distortions are more palpable for the weaker AR(1) processes ($\alpha_1 \leq 0.2$) for it is harder to estimate the AR coefficient in the first stage.

As for power figures, it appears fair to say that the HW test is a clear underperformer. The BDS test enjoys excellent power against most alternatives, easily outperforming the competition against the nonlinear MA alternative and, to a lesser extent, against the less persistent ARCH-type models. It turns out, however, that it is surprisingly powerless against the fractional AR specification. The latter is the most difficult to pinpoint among all alternatives, though our nonparametric entropy-based tests enjoy good power as long as the resulting dependence is strong enough relative to the sample size. In addition, our tests also entail relatively more power against the sign and threshold AR alternatives, especially if the serial dependence is weaker ($\alpha_1 = 0.2$).

Tables 4 and 5 document that a similar pattern arises if one employs the trimming function based on empirical quantiles, with $\kappa_2 = 0.10$. The resulting entropy-based tests entail correct size in every instance, even if applied to AR residuals. As for power, there are gains in power in every situation, with exception to the fractional AR alternative, for the sample size of 1,000 observations. This is especially the case for the nonlinear MA and ARCH-type alternatives. In contrast, it is somewhat difficult to establish strict dominance for the smaller sample size of 500 observations. For instance, the quantile-based trimming function does particularly bad for the weaker AR alternative ($\alpha_1 = 0.2$), though it brings about gains in power against nonlinear MA and ARCH-type processes.

We summarize the findings for some of the alternative trimming functions by plotting their size and power in Figure 1. We plot the results only for $q = 1/2$ and $q = 4$ in view that the power function seems monotonic with respect to the entropic index. As before, all tests exhibit correct level; the entropy-based test based on the first trimming function with $\kappa = 3/2$ entails the least size distortions. The first trimming function seems more robust to variations in κ_1 than the second is to variations in the quantile choice. Trimming out too many observations clearly hurts the performance of the entropy-based tests: e.g., there is a severe decline in power against the weaker AR alternative for $\kappa_2 = 0.3$. A comparison with Tables 2 to 5 reveals that a similar dominance pattern holds as to what concerns the two trimming functions. The quantile-based trimming helps with nonlinear MA and ARCH-type processes, but is less powerful against AR and fractional AR alternatives.

Altogether, the results are very encouraging in that our entropy-based tests seem to entail a good deal of power against several alternatives. In particular, our entropy-based statistics appears to hatch more powerful tests against nonlinear AR models such as the fractional, sign, and threshold AR processes. It remains to discuss how one should select the entropic index q so as to maximize power. Figure 1 and Tables 2 to 5 do not corroborate Tsallis's (1998) conjecture that the optimal entropic index depends on the data complexity. Power clearly increases with the entropic index for the linear and nonlinear AR processes, whereas it slightly decreases for the nonlinear MA and

ARCH-type processes. This again suggests that it would be interesting to contrive a test that integrates out the dependence on the entropic index.

5 Empirical illustration

In this section, we apply our testing procedures to check whether serial independence holds for the residuals of the predictive regression between implied and realized volatility. We thus contribute to the literature examining the information content of implied volatility (Canina and Figlewski, 1993; Christensen and Prabhala, 1998; Blair, Poon and Taylor, 2001; Bandi and Perron, 2006; Chernov, 2007) by indirectly testing for a time-varying volatility premium. In particular, we regress the monthly S&P 500 realized volatility on the VIX index of the previous month and then evaluate whether the residuals display any sort of serial dependence. The motivation is simple. Chernov (2007) argues that, to properly understand the relation between implied and realized volatility, one must account for a time-varying volatility premium. The presence of the latter will indeed contaminate the predictive regression estimates if either the contemporaneous or past volatility premium correlates with the implied volatility. Testing for serial independence of the residuals thus entails indirect evidence on whether the volatility risk premium is time varying or not.

We collect transactions data from Reuters through the Securities Industry Research Centre of Asia-Pacific (`www.sirca.org.au`) to form a time series of intraday returns on the S&P 500 index from January 3, 2000 to December 30, 2005. Our interest in the S&P 500 lies not only on the fact it is a barometer for the US economy, but also because the Chicago Board Options Exchange publishes a daily volatility index (VIX) that measures its one-month options-implied volatility. The VIX index is convenient because it is virtually free of measurement error (Doran and Ronn, 2005; Carr and Wu, 2006), controlling for the main issues raised by Hentschel (2003). To carry out the predictive regression analysis, we consider only the first daily observation of the VIX index in each month from January 2000 to December 2005.[6] In contrast, we compute the realized volatility of the S&P 500 index by summing the squares of intraday continuously compounded returns at the 15-minute frequency over the month. We employ such a frequency to balance any measurement error and/or microstructure noise that may contaminate the realized volatility estimates.

Figure 2 depicts the time series for the implied and realized volatility as well as the resulting scatter plot. As is apparent, the correlation between them is pretty strong (about 0.79), justifying to some extent the usual focus on the unbiasedness hypothesis (Canina and Figlewski, 1993; Chris-

[6] Using the last observation of the VIX index in the previous month changes very little the quantitative results.

tensen and Prabhala, 1998). To establish the informational content of the implied volatility, we run a predictive regression between implied and realized volatility, giving way to

$$\text{REALIZED VOLATILITY} = \underset{(1.4455)}{0.1823} + \underset{(0.0653)}{0.7063}\ \text{IMPLIED VOLATILITY} + \text{RESIDUAL}.$$

Despite the statistical insignificance of the intercept, the F-statistic for the null of zero intercept and unit slope indeed amounts to 108.5, with a virtually zero p-value. Figure 3 plots the actual and fitted values of the realized volatility as well as the residuals of the predictive regression. It is quite evident that the residuals display some sort of serial dependence, indicating that a time-varying volatility risk premium is very likely. Table 6 reports the results for the serial independence and correlation tests that we apply to the residuals. The Ljung-Box-Pierce tests easily reject the null of serial uncorrelation in the residuals for the residual first-order autocorrelation amounts to 0.37. In contrast, the BDS and HW tests cannot reject the null of serial independence. Although the BDS and HW tests are both completely off the mark, our nonparametric tests based on small values of the entropic index are mostly able to dismiss serial independence in the residuals at the 10% level of significance.

6 Conclusion

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, our asymptotic derivations extend 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 are suitable for assessing model specification. Second, our entropy-based tests of serial independence enjoy substantial power against both linear and nonlinear AR processes, outperforming both the BDS and HW tests. In particular, it seems that larger values of the entropic index entail more powerful tests in these instances. Third, the variation in the size and power figures resulting from changes in the entropic index and in the trimming function calls for further research on how to optimize these choices.

Finally, we employ our entropy-based tests of serial independence to investigate the informational content of the VIX index relative to the future path of the S&P 500 realized volatility. We do so by checking whether serial independence holds for the residuals of the predictive regression between implied and realized volatility. Our testing results indicate that the residuals exhibit serial dependence, which is consistent with a persistent time-varying volatility premium.

Appendix

Lemma 1. Under the null hypothesis, the following expansion holds

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

where $f_{x,y} = f_{XY}(x,y)$, $f_x = f_X(x)$, $f_y = f_Y(y)$, $h_{x,y} = \hat{f}_{XY}(x,y) - f_{XY}(x,y)$, $h_x = \int h_{x,y}\,\mathrm{d}y = \hat{f}_X(x) - f_X(x)$, $h_y = \int h_{x,y}\,\mathrm{d}x = \hat{f}_Y(y) - f_Y(y)$ and $\|\cdot\|$ denotes the sup norm on the $\mathbb{R}^2$.

Proof. Define the functional

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

where $\lambda \in [0,1]$ and

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

By Assumptions A2 and A3, the density function f_{XY} is at least thrice continuously differentiable with bounded derivatives, thus the functional above admits a third-order Taylor expansion around $\lambda = 0$ with Lagrange remainder, viz.

$$\Lambda_{f,h}(\lambda) = \Lambda_{f,h}(0) + \lambda \frac{\partial}{\partial\lambda}\Lambda_{f,h}(0) + \frac{\lambda^2}{2} \frac{\partial^2}{\partial\lambda^2}\Lambda_{f,h}(0) + \frac{\lambda^3}{6} \frac{\partial^3}{\partial\lambda^3}\Lambda_{f,h}(\lambda^*),$$

where $\lambda^* \in [0,\lambda]$. Under the null, the first and second terms of the right-hand side are zero for $h_{x,y} = \hat{f}_{XY}(x,y) - f_{XY}(x,y)$. To appreciate the latter, observe that taking the derivative of $g_{x,y}(\lambda)$ with respect to λ yields

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

Similarly,

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

which, under the null, gives way to

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

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 third and fourth terms of the expansion. Because $\frac{\partial^2}{\partial\lambda^2} g_{x,y}(\lambda) = \frac{\partial^2}{\partial\lambda^2} g_{x,y}(0) = 2h_x h_y$,

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

which, under the null hypothesis, yields

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

recovering the third term of the right-hand side. In turn, the Lagrange remainder will be cubic in $h_{x,y}$ and hence is of order $O\left(\|h_{x,y}\|^3\right)$. To complete the proof, it suffices to observe that $\Lambda_{\hat{f}} = \Lambda_{f,h}(1)$ and that the above expansion holds for all $\lambda \in [0,1]$ and, in particular, for $\lambda = 1$. This also proves the twice Hadamard differentiability of Λ_f at the true density.

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 (Bosq, 1996). The bandwidth condition in Assumption A4 then ensures that

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

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 and that the other terms will contribute only to the asymptotic bias term. Asymptotic normality follows immediately from the second-order asymptotic theory provided by Aït-Sahalia (1994) and Fan and Li (1999, Theorem 2.1), which generalize Hall's (1984) central limit theorem for degenerate U-statistics to weak dependent processes. More precisely, under Assumptions A1 to A4, it follows that

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

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

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

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. The same argument applies for the terms with lower dimensionality, though they converge at the faster rate $T\sqrt{b_T}$ and hence only their asymptotic means affect the limiting distribution of the test statistic.

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

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

Proof: By Assumptions A2 and A3, the functional

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

admits a second-order Taylor expansion around $\lambda = 0$ with Lagrange remainder

$$\Lambda^w_{f,h}(\lambda) = \Lambda^w_{f,h}(0) + \lambda\,\frac{\partial}{\partial\lambda}\Lambda^w_{f,h}(0) + \frac{\lambda^2}{2}\,\frac{\partial^2}{\partial\lambda^2}\Lambda_{f,h}(\lambda^*),$$

where $\lambda^* \in [0,\lambda]$. The first term of the right-hand side is equal to zero under the null, whereas the fact that

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

implies that, under the null,

$$\frac{\partial}{\partial\lambda}\Lambda^w_{f,h}(0) = \int w_f(x,y)\left[h_{x,y} - \frac{\partial}{\partial\lambda}g_{x,y}(0)\right]\,\mathrm{d}(x,y).$$

In view that the Lagrange remainder is quadratic in $h_{x,y}$ and hence of order $O\left(\|h_{x,y}\|^2\right)$, Λ_f is Hadamard differentiable at the true density. The result then ensues from Assumption A5(i) and the fact that $\frac{\partial}{\partial\lambda}g_{x,y}(0) = f_x h_y + f_y h_x$.

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 $\Omega = \sum_{k=-\infty}^{\infty} E[A_t A_{t+k}']$. It is straightforward to verify that, under the null of independence,

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

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

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

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

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

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 we impose are such that the functional Taylor expansion holds even when both x_{tT} and y_{tT} are double arrays. Thus, it ensues that, under $\mathbb{H}_{1,T}$ 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 $\hat{Y}_t = Y_t(\hat{\theta})$. 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 $w(\theta) = w(Y_t(\theta), Y_{t-i}(\theta))$, 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 $\theta_* \in [\theta_0, \hat{\theta}]$ and $\Lambda'_{\hat{f}}$ and $\Lambda''_{\hat{f}}$ denote the first and second order differentials with respect to θ, respectively. Let $Z_t(\theta) = (Y_t(\theta), Y_{t-1}(\theta))$ and $Z_t = (Y_t(\theta_0), Y_{t-1}(\theta_0))$. 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 $(\hat{f}_z - f_z)^2 = O_p\left(T^{-1}b_{z,T}^{-1}\right)$ and $(\hat{f}'_z - f'_z)^2 = O_p\left(T^{-1}b_{z,T}^{-3}\right)$, $\Lambda'_{\hat{f}}(\theta_0) = O_p\left(T^{-1}b_{z,T}^{-2}\right)$. 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 $\Lambda_{\hat{f}}(\hat{\theta})$ 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 $d \geq \max\{2/(s+1) - 1/2, 3/(2s+2) - 1/4\}$.

References

Aït-Sahalia, Y. (1994), The delta method for nonparametric kernel functionals, working paper, 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.

Bandi, F. and Perron, B. (2006), Long memory and the relation between implied and realized volatility, *Journal of Financial Econometrics* **4**, 636–670.

Blair, B. J., Poon, S.-H. and Taylor, S. J. (2001), Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high-frequency index returns, *Journal of Econometrics* **105**, 5–26.

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.

Canina, L. and Figlewski, S. (1993), The informational content of implied volatility, *Review of Financial Studies* **6**, 659–681.

Carr, P. and Wu, L. (2006), A tale of two indices, *Journal of Derivatives* **13**, 13–29.

Chen, X., Linton, O. and Robinson, P. M. (2001), The estimation of conditional densities, *in* M. L. Puri (ed.), *Asymptotics in Statistics and Probability*, VSP International Science Publishers, Leiden, pp. 71–84.

Chernoff, H. (1952), A measure of asymptotic efficiency for tests of hypothesis based on the sum of observations, *Annals of Mathematical Statistics* **23**, 493–507.

Chernov, M. (2007), On the role of risk premia in volatility forecasting, *forthcoming in the Journal of Business and Economic Statistics*.

Christensen, B. J. and Prabhala, N. R. (1998), The relation between implied and realized volatility, *Journal of Financial Economics* **50**, 125–150.

Cressie, N. and Read, T. R. C. (1984), Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society, Series B* **46**, 440–464.

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

Diks, C. and Manzan, S. (2002), Tests for serial independence and linearity based on correlation integrals, *Studies in Nonlinear Dynamics and Econometrics*.

Diks, C. and Panchenko, V. (2007), Rank-based entropy tests for serial independence, working paper, University of Amsterdam and University of New South Wales.

Doran, J. S. and Ronn, E. I. (2005), The bias in Black-Scholes/Black implied volatility: An analysis of equity and energy markets, *Review of Derivatives Research* **8**, 177–198.

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

Fan, Y. and Li, Q. (1999), Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing, *Journal of Nonparametric Statistics* **10**, 245–271.

Fan, Y. and Linton, O. (2003), Some higher-order theory for a consistent non-parametric model specification test, *Journal of Statistical Planning and Inference* **109**, 125–154.

Gao, J. and King, M. (2004), Model specification testing in nonparametric and semiparametric time series econometrics, *Research Report 2004/02*, School of Mathematics and Statistics, University of Western Australia.

Geweke, J. (1981), A comparison of tests of the independence of two covariance-stationary time series, *Journal of the American Statistical Association* **76**, 363–373.

Golan, A. (2002), Information and entropy econometrics: Editor's view, *Journal of Econometrics* **107**, 1–15.

Golan, A. and Perloff, J. M. (2002), Comparison of maximum entropy and higher-order entropy estimators, *Journal of Econometrics* **107**, 195–211.

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

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.

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

Härdle, W. and Mammen, E. (1993), Comparing nonparametric vs. parametric regression fits, *Annals of Statistics* **21**, 1926–1947.

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

Haugh, L. D. (1976), Checking the independence of two covariance-stationary time series: A univariate residual cross-correlation approach, *Journal of the American Statistical Association* **71**, 378–385.

Hentschel, L. (2003), Errors in implied volatility estimation, *Journal of Financial and Quantitative Analysis* **38**, 779–810.

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* **83**, 615–625.

Hong, Y. and White, H. (2005), Asymptotic distribution theory for nonparametric entropy measures of serial dependence, *Econometrica* **73**, 837–901.

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

Kočenda, E. (2001), An alternative to the BDS test: Integration across the correlation integral, *Econometric Reviews* **20**, 337–351.

Maasoumi, E. (2002), A compendium to information theory in economics and econometrics, *Econometric Reviews* **12**, 137–181.

Maasoumi, E. and Racine, J. C. (2002), Entropy and predictability of stock market returns, *Journal of Econometrics* **107**, 291–312.

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. (2000), Nonparametric misspecification testing, working paper, University of British Columbia.

Racine, J. C. and Maasoumi, E. (2007), A versatile and robust metric entropy test for time-irreversibility, and other hypotheses, *Journal of Econometrics* **138**, 547–567.

Renyi, A. (1970), *Probability Theory*, North-Holland, Amsterdam.

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.

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.

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

Figure 1
Performance of the entropy-based tests using alternative trimming functions

The first and second panels depict the rejection frequency at the 5% level for a sample size of 500 observations of the entropy-based tests of serial independence with $q = 1/2$ and $q = 4$, respectively. The third and fourth panels summarize the results for the sample size of 1,000 observations using $q = 1/2$ and $q = 4$, respectively. We consider 1,000 Monte Carlo replications and permutation/bootstrap artificial samples. We employ simple trimming functions $w(x, y) = \mathbb{1}_{\mathcal{S}}(x, y)$ that allocate weight zero to observations out of the compact set $\mathcal{S} = \mathcal{S}^X \times \mathcal{S}^Y$. We consider two types of trimming functions. The first considers a symmetric support $\mathcal{S}_1^u = \{u : |u - \bar{u}| \leq \kappa_1\, \hat{s}_u\}$ with $\kappa_1 \in \{1, 3/2, 2\}$, where $\bar{u}$ and $\hat{s}_u$ denote the sample mean and standard deviation, respectively. The second set trims observations out of a compact support $\mathcal{S}_2^u = \{u : q_u(\kappa_2) \leq u \leq q_u(1 - \kappa_2)\}$, where $q_u(\kappa_2)$ denotes the κ_2-quantile of the empirical distribution for $\kappa_2 \in \{0.10, 0.15, 0.20\}$.

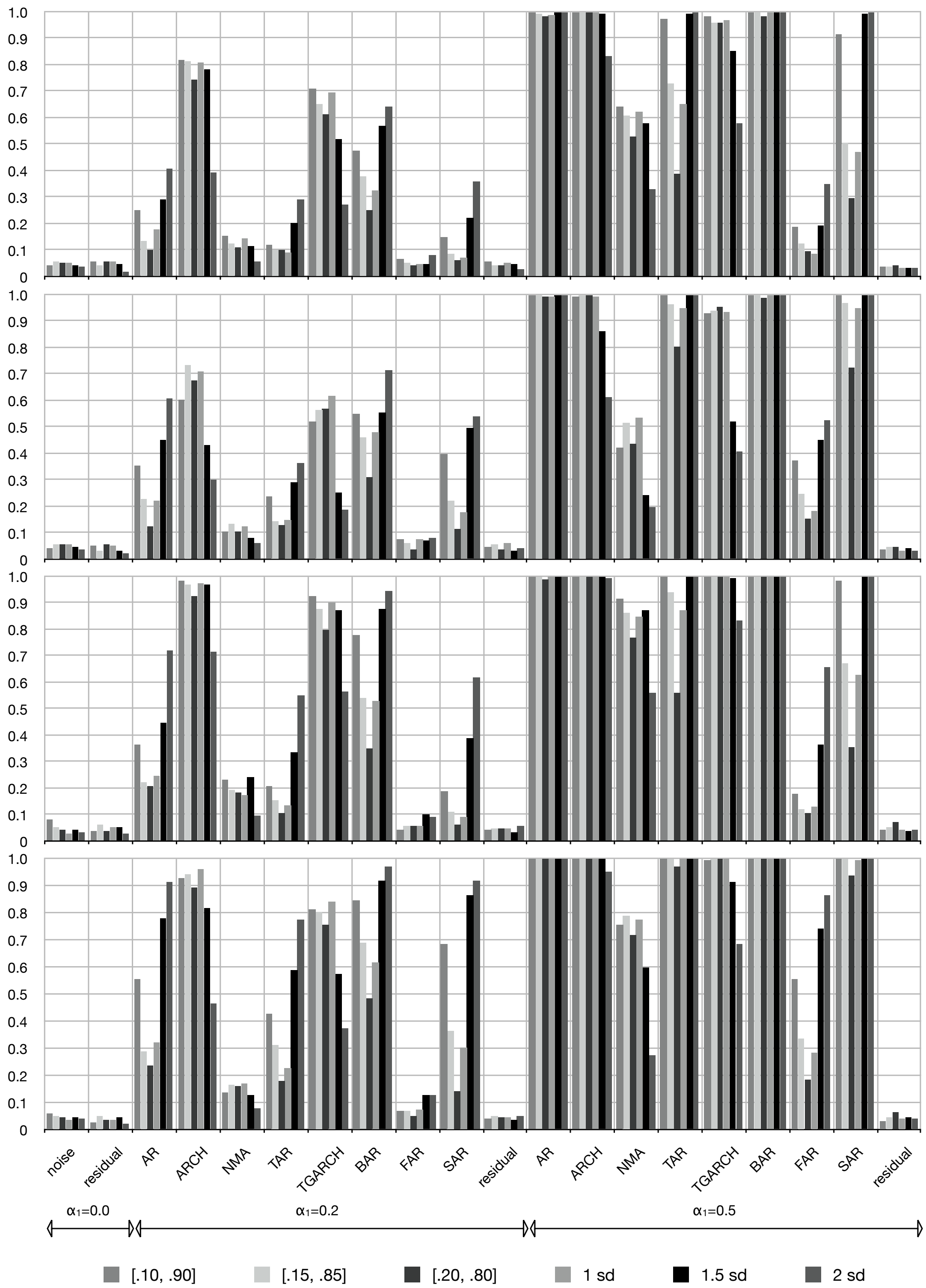

Figure 2
Time series and scatter plots of the implied and realized volatility

The left panel depicts the time series of the implied and realized volatility for the period running from January 2000 to December 2005, whereas the right panel displays the corresponding scatter plot.

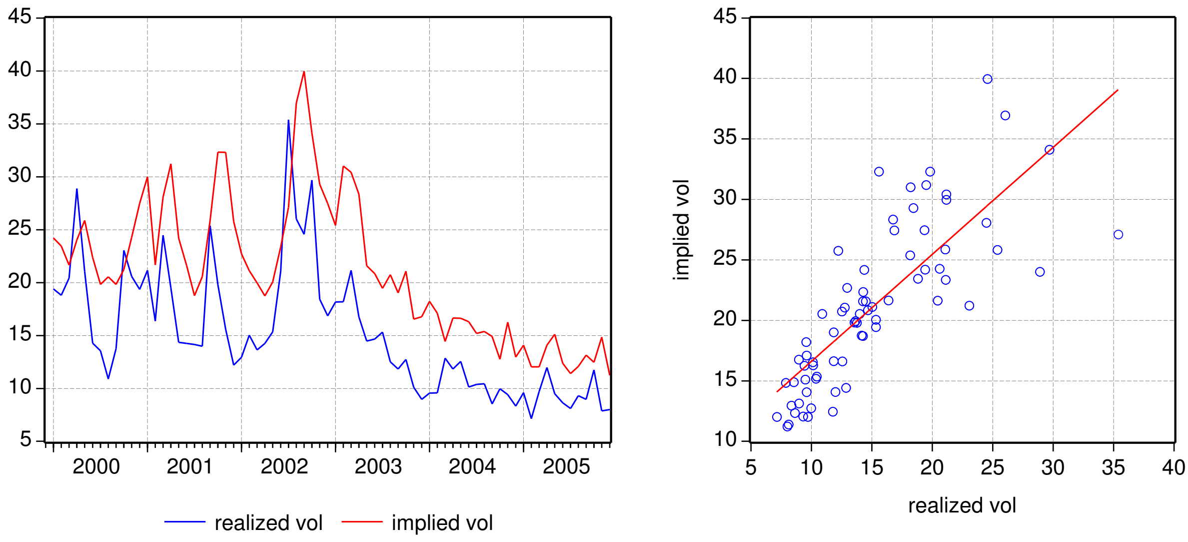

Figure 3

Results of the predictive regression between implied and realized volatility

The right vertical axis refers to the actual and predicted values of the realized volatility for a sample period running from January 2000 to December 2005, whereas the left vertical axis relates to the corresponding residuals. of the predictive regression between implied and realized volatilities.

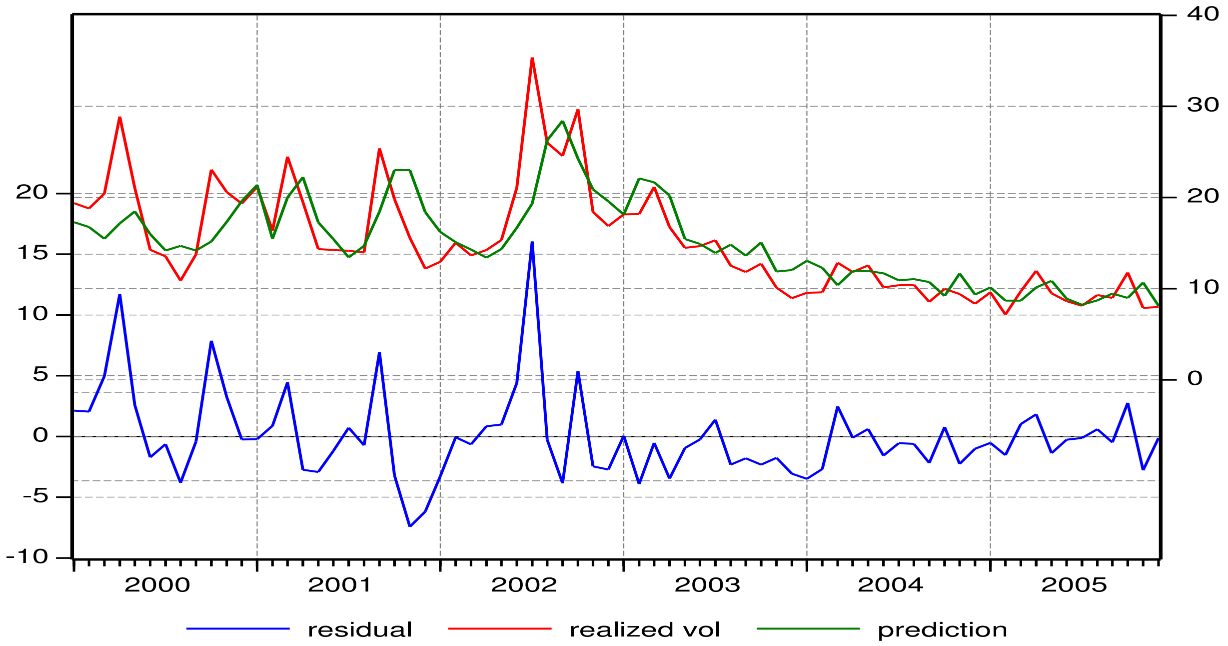

Table 1

Data generating mechanisms for the Monte Carlo simulation study

We consider two sets of parameter values for each data generating process: $(\alpha_0, \alpha_1, \alpha_2)$ is equal either to $(0.1, 0.2, 0.5)$ or to $(0.1, 0.5, 0.5)$. We vary the value of α_1 because it controls the strength of serial dependence. To mitigate any dependence on the initial values, we construct samples of 500 and 1,000 time-series observations by burning the first 500 realizations. The random noise ϵ_t is standard Gaussian regardless of the data generating process. Finally, in the last row, we compute the residuals of the AR(1) model using OLS estimates $(\hat{\alpha}_0, \hat{\alpha}_1)$ of (α_0, α_1).

data generating mechanism	model specification	initial values
Gaussian noise	$x_t = \epsilon_t$	$x_0 = \epsilon_0$
AR(1)	$x_t = \alpha_0 + \alpha_1\, x_{t-1} + \epsilon_t$	$x_0 = \epsilon_0$
ARCH(1)	$x_t = \sqrt{\alpha_0 + \alpha_1\, x_{t-1}^2}\,\epsilon_t$	$x_0 = \sqrt{\alpha_0}\,\epsilon_0$
nonlinear MA	$x_t = \alpha_1\, \epsilon_{t-1}\, \epsilon_{t-2} + \epsilon_t$	$x_0 = \epsilon_0$, $x_1 = \epsilon_1$
threshold AR(1)	$x_t = \begin{cases} \alpha_0 - \alpha_1\, x_{t-1} + \epsilon_t & \text{if } x_{t-1} \leq 1 \\ \alpha_0 + \alpha_1\, x_{t-1} + \epsilon_t & \text{if } x_{t-1} > 1 \end{cases}$	$x_0 = \epsilon_0$
threshold GARCH(1)	$x_t = h_t\, \epsilon_t$ $h_t^2 = \begin{cases} \alpha_0 + \alpha_1\, x_{t-1}^2 + \alpha_2\, h_{t-1}^2 & \text{if } \epsilon_{t-1} < 0 \\ \alpha_0 + \alpha_0\, x_{t-1}^2 + \alpha_2\, h_{t-1}^2 & \text{if } \epsilon_{t-1} \geq 0 \end{cases}$	$x_1 = \sqrt{\alpha_0}\,\epsilon_1$
bilinear AR(1)	$x_t = \alpha_0 + \alpha_1\, x_{t-1}\, \epsilon_{t-1} + \epsilon_t$	$x_0 = \epsilon_0$
fractional AR(1)	$x_t = \alpha_0 + \alpha_1\, \lvert x_{t-1}\rvert^{\alpha_2} + \epsilon_t$	$x_0 = \epsilon_0$
sign AR(1)	$x_t = \alpha_0 + \alpha_1\, \mathrm{sign}(x_{t-1}) + \epsilon_t$	$x_0 = \epsilon_0$
AR(1) residual	$x_t = \hat{\epsilon}_t(\hat{\alpha}_0, \hat{\alpha}_1) = y_t - \hat{\alpha}_0 - \hat{\alpha}_1\, y_{t-1}$	$y_0 = \epsilon_0$

Table 2

Finite-sample performance of entropy-based tests of serial independence sample size: 500 time-series observations

We compute the rejection frequency of the entropy-based tests of serial independence for $q \in \{1/2, 1, 2, 4\}$ and for the trimming function based on $\mathcal{S}_1^u = \{u: \ |u - \bar{u}| \le 2\,\hat{s}_u\}$, where $\bar{u}$ and $\hat{s}_u$ denote the sample mean and standard deviation, respectively. In addition, we also report the corresponding figures for Hong and White's (2005) test based on a Taylor expansion of the Kullback-Leibler contrast and for the BDS test with embedding dimension $m = 2$ and tuning parameter set to two standard deviations. We only consider first-order statistics in that we check exclusively for serial dependence of first order. To ensure a fair comparison, we focus exclusively on either permutation or bootstrap-based tests so as to alleviate any finite-sample size distortion. We consider a level of significance of 5%. All results rest on 1,000 Monte Carlo replications and permutation/bootstrap artificial samples.

data generating mechanism		$q = 1/2$	$q = 1$	$q = 2$	$q = 4$	HW	BDS
$\alpha_1 = 0$	Gaussian noise	0.039	0.046	0.042	0.041	0.052	0.056
	AR(1) residual	0.021	0.024	0.023	0.024	0.110	0.062
$\alpha_1 = 0.2$	AR(1)	0.409	0.473	0.555	0.612	0.127	0.462
	ARCH(1)	0.393	0.372	0.330	0.300	0.183	0.941
	nonlinear MA	0.060	0.063	0.064	0.061	0.067	0.232
	threshold AR(1)	0.291	0.341	0.387	0.368	0.125	0.114
	threshold GARCH(1,1)	0.272	0.260	0.217	0.188	0.185	0.854
	bilinear AR(1)	0.642	0.672	0.714	0.717	0.258	0.831
	fractional AR(1)	0.081	0.077	0.087	0.085	0.050	0.066
	sign AR(1)	0.359	0.439	0.529	0.543	0.142	0.154
	AR(1) residual	0.029	0.033	0.034	0.043	0.065	0.054
$\alpha_1 = 0.5$	AR(1)	1.000	1.000	1.000	1.000	0.994	1.000
	ARCH(1)	0.833	0.829	0.772	0.615	0.490	1.000
	nonlinear MA	0.333	0.302	0.264	0.199	0.174	0.849
	threshold AR(1)	1.000	1.000	1.000	1.000	0.969	0.675
	threshold GARCH(1,1)	0.579	0.561	0.511	0.409	0.288	0.993
	bilinear AR(1)	1.000	1.000	1.000	1.000	0.896	1.000
	fractional AR(1)	0.353	0.432	0.507	0.527	0.132	0.155
	sign AR(1)	0.998	0.999	1.000	1.000	0.950	0.823
	AR(1) residual	0.033	0.034	0.032	0.034	0.082	0.040

Table 3

Finite-sample performance of entropy-based tests of serial independence sample size: 1,000 time-series observations

We compute the rejection frequency of the entropy-based tests of serial independence for $q \in \{1/2, 1, 2, 4\}$ and for the trimming function based on $\mathcal{S}_1^u = \{u : |u - \bar{u}| \leq 2\,\hat{s}_u\}$, where $\bar{u}$ and $\hat{s}_u$ denote the sample mean and standard deviation, respectively. In addition, we also report the corresponding figures for Hong and White's (2005) test based on a Taylor expansion of the Kullback-Leibler contrast and for the BDS test with embedding dimension $m = 2$ and tuning parameter set to two standard deviations. We only consider first-order statistics in that we check exclusively for serial dependence of first order. To ensure a fair comparison, we focus exclusively on either permutation or bootstrap-based tests so as to alleviate any finite-sample size distortion. We consider a level of significance of 5%. All results rest on 1,000 Monte Carlo replications and permutation/bootstrap artificial samples.

data generating mechanism		$q = 1/2$	$q = 1$	$q = 2$	$q = 4$	HW	BDS
$\alpha_1 = 0$	Gaussian noise	0.034	0.035	0.043	0.044	0.048	0.060
	AR(1) residual	0.028	0.023	0.026	0.025	0.074	0.041
$\alpha_1 = 0.2$	AR(1)	0.724	0.837	0.890	0.914	0.288	0.692
	ARCH(1)	0.717	0.682	0.580	0.468	0.369	0.997
	nonlinear MA	0.099	0.085	0.079	0.079	0.091	0.309
	threshold AR(1)	0.550	0.638	0.737	0.775	0.211	0.158
	threshold GARCH(1,1)	0.564	0.551	0.487	0.376	0.292	0.983
	bilinear AR(1)	0.947	0.962	0.970	0.973	0.576	0.964
	fractional AR(1)	0.094	0.110	0.132	0.130	0.082	0.067
	sign AR(1)	0.619	0.759	0.867	0.920	0.260	0.164
	AR(1) residual	0.057	0.054	0.053	0.053	0.075	0.052
$\alpha_1 = 0.5$	AR(1)	1.000	1.000	1.000	1.000	1.000	1.000
	ARCH(1)	0.994	0.994	0.991	0.953	0.812	1.000
	nonlinear MA	0.562	0.484	0.363	0.274	0.310	0.976
	threshold AR(1)	1.000	1.000	1.000	1.000	1.000	0.900
	threshold GARCH(1,1)	0.850	0.835	0.791	0.686	0.516	1.000
	bilinear AR(1)	1.000	1.000	1.000	1.000	0.998	1.000
	fractional AR(1)	0.660	0.764	0.829	0.867	0.213	0.193
	sign AR(1)	1.000	1.000	1.000	1.000	1.000	0.959
	AR(1) residual	0.044	0.045	0.041	0.043	0.082	0.055

Table 4

Finite-sample performance of entropy-based tests of serial independence sample size: 500 time-series observations

We compute the rejection frequency of the entropy-based tests of serial independence for $q \in \{1/2, 1, 2, 4\}$ and for the trimming function based on the compact support $\mathcal{S}_2^u = \{u : q_u(0.10) \leq u \leq q_u(0.90)\}$, where $q_u(\kappa)$ denotes the κ-quantile of the empirical distribution. In addition, we also report the corresponding figures for Hong and White's (2005) test based on a Taylor expansion of the Kullback-Leibler contrast and for the BDS test with embedding dimension $m = 2$ and tuning parameter set to two standard deviations. We only consider first-order statistics in that we check exclusively for serial dependence of first order. To ensure a fair comparison, we focus exclusively on either permutation or bootstrap-based tests so as to alleviate any finite-sample size distortion. We consider a level of significance of 5%. All results rest on 1,000 Monte Carlo replications and permutation/bootstrap artificial samples.

data generating mechanism		$q = 1/2$	$q = 1$	$q = 2$	$q = 4$	HW	BDS
$\alpha_1 = 0$	Gaussian noise	0.046	0.047	0.044	0.042	0.052	0.056
	AR(1) residual	0.058	0.058	0.055	0.054	0.110	0.062
$\alpha_1 = 0.2$	AR(1)	0.255	0.270	0.316	0.357	0.127	0.462
	ARCH(1)	0.819	0.797	0.755	0.604	0.183	0.941
	nonlinear MA	0.158	0.145	0.146	0.108	0.067	0.232
	threshold AR(1)	0.121	0.139	0.201	0.240	0.125	0.114
	threshold GARCH(1,1)	0.713	0.687	0.638	0.522	0.185	0.854
	bilinear AR(1)	0.478	0.518	0.547	0.553	0.258	0.831
	fractional AR(1)	0.070	0.071	0.074	0.079	0.050	0.066
	sign AR(1)	0.151	0.205	0.303	0.402	0.142	0.154
	AR(1) residual	0.057	0.057	0.058	0.047	0.065	0.054
$\alpha_1 = 0.5$	AR(1)	0.999	1.000	1.000	1.000	0.994	1.000
	ARCH(1)	1.000	1.000	1.000	0.996	0.490	1.000
	nonlinear MA	0.643	0.625	0.554	0.422	0.174	0.849
	threshold AR(1)	0.976	0.991	0.999	1.000	0.969	0.675
	threshold GARCH(1,1)	0.985	0.976	0.966	0.930	0.288	0.993
	bilinear AR(1)	1.000	1.000	1.000	1.000	0.896	1.000
	fractional AR(1)	0.189	0.226	0.288	0.378	0.132	0.155
	sign AR(1)	0.917	0.969	0.995	0.999	0.950	0.823
	AR(1) residual	0.037	0.036	0.037	0.041	0.082	0.040

Table 5

Finite-sample performance of entropy-based tests of serial independence sample size: 1,000 time-series observations

We compute the rejection frequency of the entropy-based tests of serial independence for $q \in \{1/2, 1, 2, 4\}$ and for the trimming function based on the compact support $\mathcal{S}_2^u = \{u : q_u(0.10) \leq u \leq q_u(0.90)\}$, where $q_u(\kappa_2)$ denotes the κ_2-quantile of the empirical distribution. In addition, we also report the corresponding figures for Hong and White's (2005) test based on a Taylor expansion of the Kullback-Leibler contrast and for the BDS test with embedding dimension $m = 2$ and tuning parameter set to two standard deviations. We only consider first-order statistics in that we check exclusively for serial dependence of first order. To ensure a fair comparison, we focus exclusively on either permutation or bootstrap-based tests so as to alleviate any finite-sample size distortion. We consider a level of significance of 5%. All results rest on 1,000 Monte Carlo replications and permutation/bootstrap artificial samples.

data generating mechanism		$q = 1/2$	$q = 1$	$q = 2$	$q = 4$	HW	BDS
$\alpha_1 = 0$	Gaussian noise	0.083	0.074	0.062	0.062	0.048	0.060
	AR(1) residual	0.041	0.038	0.031	0.030	0.074	0.041
$\alpha_1 = 0.2$	AR(1)	0.368	0.415	0.482	0.558	0.288	0.692
	ARCH(1)	0.983	0.978	0.971	0.927	0.369	0.997
	nonlinear MA	0.232	0.225	0.205	0.136	0.091	0.309
	threshold AR(1)	0.209	0.242	0.336	0.429	0.211	0.158
	threshold GARCH(1,1)	0.925	0.920	0.903	0.815	0.292	0.983
	bilinear AR(1)	0.780	0.806	0.835	0.846	0.576	0.964
	fractional AR(1)	0.044	0.049	0.057	0.073	0.082	0.067
	sign AR(1)	0.190	0.296	0.472	0.684	0.260	0.164
	AR(1) residual	0.045	0.043	0.044	0.044	0.075	0.052
$\alpha_1 = 0.5$	AR(1)	1.000	1.000	1.000	1.000	1.000	1.000
	ARCH(1)	1.000	1.000	1.000	1.000	0.812	1.000
	nonlinear MA	0.917	0.897	0.864	0.755	0.310	0.976
	threshold AR(1)	0.999	1.000	1.000	1.000	1.000	0.900
	threshold GARCH(1,1)	1.000	1.000	1.000	0.997	0.516	1.000
	bilinear AR(1)	1.000	1.000	1.000	1.000	0.998	1.000
	fractional AR(1)	0.180	0.258	0.381	0.555	0.213	0.193
	sign AR(1)	0.983	0.999	1.000	1.000	1.000	0.959
	AR(1) residual	0.045	0.045	0.036	0.033	0.082	0.055

Table 6

Testing results for the residuals of the predictive regression between implied and realized volatility

We report the p-values of the autocorrelation and serial independence tests that we employ to examine the residuals of the predictive regression between implied and realized volatility. The sample period ranges from January 2000 to December 2005, adding up to 72 monthly observations. To test for serial correlation, we compute the Ljung-Box-Pierce statistic of first order, which essentially boils down to the squared value of the first-order autocorrelation of the residuals. To look into serial independence, we employ our nonparametric entropy-based tests using simple trimming functions $w(x, y) = \mathbb{1}_{\mathcal{S}}(x, y)$ that allocate weight zero to observations out of the compact set $\mathcal{S} = \mathcal{S}^X \times \mathcal{S}^Y$. We consider two types of trimming functions. The first considers a symmetric support $\mathcal{S}_1^u = \{u : |u - \bar{u}| \leq \kappa_1 \hat{s}_u\}$ with $\kappa_1 \in \{1, 3/2, 2\}$, where $\bar{u}$ and $\hat{s}_u$ denote the sample mean and standard deviation, respectively. The second set trims observations out of a compact support $\mathcal{S}_2^u = \{u : q_u(\kappa_2) \leq u \leq q_u(1 - \kappa_2)\}$, where $q_u(\kappa_2)$ denotes the κ_2-quantile of the empirical distribution for $\kappa_2 \in \{0.10, 0.15, 0.20\}$. In addition, we also report the p-values of the Hong-White test based on a Taylor expansion of the Kullback-Leibler contrast and of the BDS test with embedding dimension $m = 2$ and tuning parameter set to κ standard deviations. Due to the small number of observations, we compute more precise critical values for the above tests by forming $B = 9,999$ artificial bootstrap samples of the residuals.

test of serial correlation/dependence			(bootstrap-based) p-value		
Ljung-Box-Pierce statistic					0.016
Hong-White statistic					0.140
BDS statistic			$\kappa = 1/2$	$\kappa = 1$	$\kappa = 2$
			0.240	0.236	0.225
entropy-based tests		$q = 1/2$	$q = 1$	$q = 2$	$q = 4$
first trimming	$\kappa_1 = 1$	0.094	0.132	0.227	0.436
	$\kappa_1 = 3/2$	0.049	0.093	0.227	0.544
	$\kappa_1 = 2$	0.161	0.258	0.473	0.782
second trimming	$\kappa_2 = 0.10$	0.101	0.143	0.243	0.465
	$\kappa_2 = 0.15$	0.048	0.065	0.116	0.246
	$\kappa_2 = 0.20$	0.038	0.048	0.069	0.125